

Attach to Paper

Access DB# _____

SEARCH REQUEST FORM

Scientific and Technical Information Center

Requester's Full Name: _____ Examiner #: _____ Date: _____
Art Unit: _____ Phone Number 30 _____ Serial Number: _____
Mail Box and Bldg/Room Location: Room 111 Results Format Preferred (circle): PAPER DISK E-MAIL

If more than one search is submitted, please prioritize searches in order of need.

Please provide a detailed statement of the search topic, and describe as specifically as possible the subject matter to be searched. Include the elected species or structures, keywords, synonyms, acronyms, and registry numbers, and combine with the concept or utility of the invention. Define any terms that may have a special meaning. Give examples or relevant citations, authors, etc, if known. Please attach a copy of the cover sheet, pertinent claims, and abstract.

Title of Invention: _____

Inventors (please provide full names): _____

Earliest Priority Filing Date: _____

**For Sequence Searches Only* Please include all pertinent information (parent, child, divisional, or issued patent numbers) along with the appropriate serial number.*

STAFF USE ONLY

	Type of Search	Vendors and cost where applicable
Searcher: _____	NA Sequence (#) _____	STN _____
Searcher Phone #: _____	AA Sequence (#) _____	Dialog _____
Searcher Location: _____	Structure (#) _____	Questel/Orbit _____
Date Searcher Picked Up: _____	Bibliographic _____	Dr.Link _____
Date Completed: _____	Litigation _____	Lexis/Nexis _____
Searcher Prep & Review Time: _____	Fulltext _____	Sequence Systems _____
Clerical Prep Time: _____	Patent Family _____	WWW/Internet _____
Online Time: _____	Other _____	Other (specify) _____

Query Match	1.08; Score 65.6; DB 1; Length 7218
-------------	-------------------------------------


```

Oy 3869 ACGTGGCTAAAGATTATGGAAGAATAAATCTCCTCACTTAATCATAGATGCAT 3928
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 553 AATGTCGAATTAGCTAGTGAGAGAGTAAAGAAAAAATTTCTGACCTATTAGAGAGGA 592
Oy 3929 ACTGAACCTTGTACAGACCTCGGAAAATGATAGCAATTCACAACGATCCGACATTCG 3988
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 593 AATCACTTAAGCTGAAGTGTGATGATTAATTAATAATTTAGAGAAGCCGAATATMAAG 652
Oy 3969 AAAAATGGAATATCACCAAGTATTAAATGAGAAAAATCTGAATATTCAAAAACACATT 4048
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 653 GAAATATACTTATTAGTAATATGGAAGAACCAAAGAAAAATATATTGACAAATTATTA 712
Oy 4049 GAAAGTACTCTGCTGATPAAGATGAGATGGATCAATTTGAAACGTTATGTTGANAAT 4108
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 713 AATATATTTGGACAAATATTCGAAAAAACMGAAAGTGTATCAGAAATGTACACAGTCAGT 772
Oy 4109 GAGAAATGTATTAAATGACGATGACATTCCTCAACCTAATCCATACCAGATGAAATTAAT 4168
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 773 GATGAACCTTTTTAATGATTTATTAATAATAGTGTATGATGAGAAAGTAAAAAGAAAT 832
Oy 4169 GATGTCGATGATGAGACGTTTGATTAATATATCATATGACATGTCACAGACGAAAGAAA 4228
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 833 ATTTTGGAGGAAGCTCAAGTTAAAGACGATATTTTATAGTTAGTAAAAAGTGTCAA 892
Oy 4229 TACACATTTGAGAAAGAAATTAACGAAAAAATCTGCTACTAAACATTCAGTTGA 4284
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 893 CAAGACACACACACCAATGTTGAAGAAAAAAGTTGAAGAAAGTGTAGAGAAATGA 948

RESULT 6
US-09-410-464-9
; Sequence 9, Application US/09410464
; Patent No. 6395892
; GENERAL INFORMATION:
; APPLICANT: Strauss et al.
; TITLE OF INVENTION: floral homeotic genes for manipulation of flowering in
; TITLE OF INVENTION: poplar and other plant species.
; FILE REFERENCE: 53375
; CURRENT APPLICATION NUMBER: US/09/410,464
; EARLIER FILING DATE: 1999-10-01
; EARLIER APPLICATION NUMBER: 09/287,700

```

EARLIER FILING DATE: 1999-04-06
 EARLIER APPLICATION NUMBER: 60/080,851
 EARLIER FILING DATE: 1998-04-06
 NUMBER OF SEQ ID NOS: 24
 SOFTWARE: PatentIn Ver. 2.0
 SEQ ID NO 9
 LENGTH: 11485
 TYPE: DNA
 ORGANISM: Populus balsamifera subsp. trichocarpa
 US-09-410-464-9

Query Match 0.98; Score 55.6; DB 4; Length 11485;
 Best Local Similarity 44.68; Pred. No. 0.0034;
 Matches 348; Conservative 0; Mismatches 424; Indels 8; Gaps 3;

4468 CACAAATTCAGTATATGATCCGGATGATTAAGTCATGATGCTGAAGTACAGAA 4527
 130 CTATTAAGAGGCAATCTGATCGCTCGGCAACAGCATATGATTAAGATTTCTGC 169
 4528 ATTATGATCAAGATGTTTACGAGAGATTCACATCCACCGGTGTAAGCTATATC 4587
 190 TTTCATTAAGACAGATCTGGATCTGTTCTCTACCTCCCGTAAGAGTGTGG 249
 4588 TATGGGTGGGTACATCTGAGAAATGATCTCTCAAGAGTGTGTTGGAAATCAG 4647
 250 TTGTCATGGGTGATTAAGATCAAGACTAATCT--GATGGGTCTATTGACATCAA 306
 4648 TTGTTGTCATGACGACAGACAAAGAAATGATGATGACCTTTTATGTTAG 4707
 307 AGCTAGGCTGTTGCAAAAGATCTCTCAACATTAAGTATGACATATGAGAAAT 366
 4708 TTACCTGTTATAGATCTTGACTATTAAGTATTAACATTAATAGTTGATTAAG 4767
 367 TGCCCGGGTCCAAAAAGACTACTATCTGACTTATGTCGACTGATGATCGTCA 426
 4768 AATGACATTCACATTAAGTACGTCGAGTGGCGTATCTAAATGCTCTATTACATTC 4827
 427 GTGCAATTTCTCAGCTGATGTTAAATGCTCTTGAATGAGATCTTCAAGAGA 486
 4828 AATCCATTTATGCTTCTCCCTAAATAGTACCTTTGAAGAAACCATGTTGCT 4887
 487 AGTTATGTCGACCTCCCTCGATTTATGATGACGC--TGATATGTTTGAACCT 544
 4888 ATTGAAGCTTCTGCTATGAGTTAAACAGTGGGTTGATGATGATGATCAACAA 4947
 545 AAGAAAGCATTAATTAATGCTCAAAAGACCCCGTCTGTTGAGAAATCTC 604
 4948 AAGATTTGGAAGACATGTTTACTCAAGTTTACACATGATGATGATTTTACAT 5007
 605 TATGTGATCTGCTCTGTCGATTTTCTAGCATGATGATGCTCTCTTTT--AT 661
 5008 TGAATATGAGAGGATGATTAATTAAGTTTATGATGATGATGATGATGATGAT 5067
 662 TAAGTCAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 721
 5068 TGAAGTTCACAAAAGTATGATTAATTTTGTGATCAATGAGATATTTTGAAGT 5127
 722 TGGGATGACATGATGATGATGATGATGATGATGATGATGATGATGATGAT 781
 5128 TAAAGTGTGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 5187
 782 GAAAGATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 841
 5188 TTATATTTTATCAAGAAATTTCTCAAGAAATTTTATTAAGATTTCAACATAGAT 5247
 842 TTACCTTCTCTCAGTCAAGAAATTTGTCAGATATTTCTGAGACGATGATGATGAT 901

RESULT 7
 US-08-771-602D-1
 Sequence 1, Application us/08771602D
 Patent No. 5976795
 GENERAL INFORMATION:

APPLICANT: Voytas, Daniel F.
 APPLICANT: Zou, Sige
 TITLE OF INVENTION: Retrotransposon and Methods
 NUMBER OF SEQUENCES: 51
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Greenlee, Winner and Sullivan, P.C.
 STREET: 5370 Manhattan Circle, Suite 201
 CITY: Boulder
 STATE: Colorado
 COUNTRY: USA
 ZIP: 80303
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/771,602D
 FILING DATE: 20-DEC-1996
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 60/010,869
 FILING DATE: 31-JAN-1996
 ATTORNEY/AGENT INFORMATION:
 NAME: Ferber, Donna M.
 REGISTRATION NUMBER: 33,878
 REFERENCE/DOCKET NUMBER: 8-96
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (303) 499-8080
 TELEFAX: (303) 499-8089
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 6660 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: not relevant
 MOLECULE TYPE: DNA (genomic)
 HYPOTHETICAL: NO
 ANTI-SENSE: NO
 ORIGINAL SOURCE:
 ORGANISM: Saccharomyces paradoxus
 STRAIN: NRRL Y-17217
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 1441..6321
 FEATURE:
 NAME/KEY: misc-feature
 LOCATION: 1228..6602
 OTHER INFORMATION: /function="retrotransposon"
 OTHER INFORMATION: /product="Ty5-6p"
 FEATURE:
 NAME/KEY: misc-feature
 LOCATION: 1228..1478
 OTHER INFORMATION: /function="5' LTR of Ty5-6p"
 FEATURE:
 NAME/KEY: misc-feature
 LOCATION: 6352..6602
 OTHER INFORMATION: /function="3' LTR of Ty5-6p"
 FEATURE:
 NAME/KEY: misc-feature
 LOCATION: 2852..4827
 OTHER INFORMATION: /function="integrase region coding"
 OTHER INFORMATION: region of Ty5-6p"
 US-08-771-602D-1

Query Match 0.88; Score 54.6; DB 2; Length 6660;
 Best Local Similarity 45.28; Pred. No. 0.0077;
 Matches 241; Conservative 0; Mismatches 289; Indels 3; Gaps 1;
 4661 GCCACAGACAAAGAAATTTGATTAATGATGATGATGATGATGATGATGATGAT 4720
 4981 GCCAAGGACATATCAAGAGCTGATGATGATGATGATGATGATGATGATGATGAT 5040

[illegible]

	Query Match	Similarity	Score	DB	Length	
	Best Local	59.9%	Pred.	No. 0.0067	Mismatches	
	Matches	Conservative	0;	Mismatches	Indels Gaps	
QY	1211	TTCACGACACAGNAAACC	AAGAAAAGAA	CACAGACAGAACCTC	ATCGAACAATCTCA	1270
Db	3121	TTTGCATGCCAAAAA	AAAAAAAAAAAAAAAA	AAAAAAAAAAAAAAAA	AAAAAAAAAAAAAAAA	3180
QY	1271	AAGAAGACTGTAATCAA	GAAAGAAATA	GAAACATCCAAAT	CAGATTAACGATAA	1330
Db	3181	AAAAAAAAAAAAAAAA	AAAAAAAAAAAAAAAA	AAAAAAAAAAAAAAAA	AAAAAAAAAAAAAAAA	3240
QY	1331	GGTGAAAAAGAAAAAG	AAAAAAACCTTC			1362
Db	3241	AAAAAAAAAAAAAAAA	AAAAAAAAAACTTC			3272

RESULT 10
US-09-797-906-1
; Sequence 1, Application US/09797906
; Patent No. 6329188
; GENERAL INFORMATION
; APPLICANT: ZIANGHE YAN, Karen A. KETCHUM, Valentina DIFRANCESCO, Ellen M. BEASLEY

05-09-150-741-1	
Query Match	0.88; Score 53.2; DB 4; Length 8920;

2.

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd

OM nucleic - nucleic search, using sw model

Run on: August 24, 2001, 22:54:05 ; Search time 356.12 Seconds

11330.158 Million cell updates/sec

Title: US-09-430-590E-3
Perfect score: 6436

Sequence: 1 tgttggttgtcactatt.....agaaagttatatattccatca 6426

Scoring table: IDENTITY_NUC

Searched: 730101 seqs, 313950809 residues

Total number of hits satisfying chosen parameters: 1460202

Maximum DB seq length:	200000000000
Minimum DB seq length:	0

Post-processing: Minimum Match 08

Listing first 45 summaries

Database :

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	6424.4	100.0	6426	21	AAAS7920	Candida albicans F
2	388.8	6.1	392	21	AAAS7936	Candida albicans s
3	388.8	6.1	392	21	AAAS7923	Candida albicans s
4	388.8	6.1	392	21	AAAS7927	Candida albicans s
5	388.8	6.1	392	21	AAAS7928	Candida albicans s
6	388.8	6.1	392	21	AAAS7931	Candida albicans s
7	388.8	6.1	392	21	AAAS7932	Candida albicans s
8	387.2	6.0	392	21	AAAS7925	Candida albicans s
9	387.2	6.0	392	21	AAAS7929	Candida albicans s
10	387.2	6.0	392	21	AAAS7933	Candida albicans s
11	385.6	6.0	392	21	AAAS7930	Candida albicans s

12	384	6.0	392	21	AAAS7934	Candida albicans s
13	384	6.0	392	21	AAAS7934	Candida albicans s
14	382.4	6.0	392	21	AAAS7934	Candida albicans s
15	380.8	5.9	392	21	AAAS7926	Candida albicans s
16	328	5.1	1385	21	AAAS7989	Candida albicans s
17	280	4.4	1249	21	AAAS7947	1249 bp Candida al
18	280	4.4	1483	21	AAAS7990	1483 bp Candida al
19	280	4.4	5611	21	AAAS7948	5611 bp Candida al
20	275.2	4.3	879	21	AAAS7991	879 bp Candida alb
21	264.4	4.1	454	21	AAAS7967	Candida albicans m
22	255.6	4.0	974	21	AAAS7992	Candida albicans m
23	252.8	3.9	1308	21	AAAS7949	1308 bp Candida alb
24	249.8	3.9	469	21	AAAS7994	469 bp Candida alb
25	224.6	3.5	9850	21	AAAS8018	9850 bp Candida al
26	224.6	3.5	3604	21	AAAS7946	3604 bp Candida al
27	190	3.0	690	21	AAAS7951	690 bp Candida alb
28	174.4	2.7	936	22	AAAS8252	Oligonucleotide D1
29	174.4	2.7	936	22	AAAS8254	Oligonucleotide D1
30	174.4	2.7	936	22	AAAS8257	Oligonucleotide D1
31	174.4	2.7	936	22	AAAS8259	Oligonucleotide D2
32	174.4	2.7	936	22	AAAS8259	Oligonucleotide D2
33	174.4	2.7	936	22	AAAS8255	Oligonucleotide D1
34	172.2	2.7	936	22	AAAS8254	Oligonucleotide D1
35	172.2	2.7	936	22	AAAS8257	Oligonucleotide D1
36	172.2	2.7	936	22	AAAS8259	Oligonucleotide D1
37	172.2	2.7	936	22	AAAS8259	Oligonucleotide D2
38	172.2	2.7	936	22	AAAS8259	Oligonucleotide D2
39	172.2	2.7	936	22	AAAS8255	Oligonucleotide D2
40	170.4	2.7	1912	21	AAAS7952	1912 bp Candida al
41	164	2.6	1672	21	AAAS7950	1672 bp Candida al
42	130	2.0	138	21	AAAS7937	Candida albicans s
43	104.2	1.6	770	21	AAAS7964	770 bp Candida alb
44	104.2	1.6	951	21	AAAS8017	951 bp Candida alb
45	100.6	1.6	1348	21	AAAS7983	1348 bp Candida al

ALIGNMENTS

RESULT	1
AAA57920	
ID	AAA57920 standard; DNA; 6426 BP.
XX	
AC	AAA57920;
XX	
DT	11-OCT-2000 (first entry)
XX	
DE	Candida albicans pCal retrotransposon.
XX	
KW	Retrotransposon; pCal; TCa2; Ty1; copia; long terminal repeat; LTR;
KW	gag gene; group antigen; polyprotein; pol; aspartate protease; integrase
KW	reverse transcriptase; RNaseH; pseudonot; readthrough translation;
KM	stop codon suppression; gene delivery; gene therapy vector;
XX	genetic vaccine composition; immunogenic; transgenic animal; ds.
OS	Candida albicans strain hog1042.
XX	
XX	
Key	Location/Qualifiers
FH	1..280
FT	/*tag= a
FT	1..6
FT	/*tag= b
FT	/rpt_type= INVERTED
FT	/note= "Imperfect 6 bp repeat"
FT	120..125
FT	/*tag= c
FT	/standard name= "TATA box"
FT	185..190
FT	/*tag= d
FT	/standard name= "TATA box"
FT	201..206
FT	/*tag= e
FT	275..280
FT	repeat_unit

```

FT      /tag= f
FT      /rpt_type= INVERTED
FT      /note= "Imperfect 6 bp repeat"
FT      281..291
FT      /tag= g
FT      /note= "this site in the corresponding RNA is a primer
FT      binding site for reverse transcriptase-mediated
FT      minus-strand DNA synthesis"
FT
FT      CDS
FT      498..1372
FT      /tag= h
FT      /product= "Gag (group antigen) protein"
FT      /codon= (seq:"ctg", aa:Ser)
FT      /note= "5 serine residues in the gag protein are encoded
FT      by this non-standard Ser codon"
FT      498..6103
FT      /tag= i
FT      /product= "Gag (group antigen)-pol (polyprotein)
FT      readthrough protein"
FT      /note= "The gag-pol readthrough protein is the result of
FT      stop codon suppression mediated by pseudoknot
FT      formation in the mRNA"
FT      /codon= (seq:"ctg", aa:Ser)
FT      /note= "12 serine residues in the gag-pol readthrough
FT      protein are encoded by this non-standard Ser
FT      codon"
FT      /transl_except= (pos:1367..1372, aa:Glu)
FT      1381..1455
FT      /tag= j
FT      /note= "The RNA corresponding to this region forms a
FT      pseudoknot, allowing gag ORF stop codon
FT      suppression and translation of the gag-pol
FT      readthrough protein"
FT      1373..6100
FT      /tag= k
FT      /product= "Pol (polyprotein), comprising aspartate
FT      protease, integrase, reverse transcriptase
FT      (RT) and RNaseH"
FT      /codon= (seq:"ctg", aa:Ser)
FT      /note= "7 serine residues in the pol protein are encoded
FT      by this non-standard Ser codon"
FT      3455..3465
FT      /tag= l
FT      /note= "Polypurine tract 2 (PPT2)"
FT      /complement= (3455..3465)
FT      /tag= m
FT      /note= "this site in the corresponding RNA is a primer
FT      binding site for reverse transcriptase-mediated
FT      plus-strand DNA synthesis"
FT      6136..6146
FT      /tag= n
FT      /note= "Polypurine tract 1 (PPT1)"
FT      /complement= (6136..6146)
FT      /tag= o
FT      /note= "this site in the corresponding RNA is a primer
FT      binding site for reverse transcriptase-mediated
FT      plus-strand DNA synthesis"
FT
FT      primer_bind
FT      6147..6426
FT      /tag= p
FT      repeat_unit
FT      6147..6152
FT      /tag= q
FT      /rpt_type= INVERTED
FT      /note= "Imperfect 6 bp repeat"
FT      6266..6271
FT      /tag= r
FT      /standard name= "TATA box"
FT      6331..6336
FT      /tag= s
FT      /standard name= "TATA box"
FT      6347..6351
FT      /tag= t
FT      6421..6426
FT      /tag= u
FT      /rpt_type= INVERTED

```

```

FT      /note= "Imperfect 6 bp repeat"
XX      WO200026397-A1.
XX      11-MAY-2000.
XX      01-NOV-1999; 99WO-N200179.
XX      30-OCT-1998; 98CA-2249046.
XX      30-OCT-1998; 98US-0106342.
XX      (JANC ) JANSEN PHARM NV.
XX      Luyten WHML, De Backer MD, Nelissen BJM, Poulter RTM;
XX      WPI: 2000-365640/31.
XX      P-PSDB; AAB03126, AAB03127, AAB03128.
XX
XX      Novel retrotransposon expression vectors useful for expressing an
XX      antigen, epitope or therapeutic agent, or detecting genes or the
XX      presence of Candida in a sample -
XX
XX      Claim 12; Fig 2B; 204pp; English.
XX
XX      The invention relates to novel retrotransposons from the yeast Candida
XX      albicans which have a copy number of 40-150, preferably 50-100 copies
XX      per genome. In particular, the invention relates to the novel C.
XX      albicans Ty1/copia retrotransposon pcal (AAB57920), and to the integrated
XX      form of this retrotransposon, designated rca2, and to the novel C.
XX      albicans retrotransposons 1-28. pcal was initially isolated from C.
XX      albicans HOG1042 and has a copy number of 50-100 copies per cell. It
XX      comprises identical 280 bp long terminal repeats (LTRs) and two open
XX      reading frames (ORFs). The first ORF encodes a gag (group antigen)
XX      protein, and the second ORF encodes a polyprotein (pol) consisting of an
XX      aspartate protease, integrase, reverse transcriptase (RT) and RNaseH. The
XX      gag and pol ORFs of pcal are in the same reading frame, separated only by
XX      a termination codon (TGA). Translation of the pol ORF occurs through
XX      the occasional readthrough suppression of the stop codon, which is
XX      mediated by the formation of a pseudoknot within the gag-pol mRNA.
XX      The retrotransposons of the invention can be used as vectors for in
XX      vitro or in vivo transformation and expression. They can thus be used
XX      for the delivery and expression of a therapeutic, immunological or
XX      immunogenic molecule (e.g., an antigen) and may also be used for
XX      eliciting an immunological response in a host organism. They are
XX      therefore useful in genetic vaccine compositions and for gene therapy,
XX      particularly where the use of retroviral vectors is unsate or
XX      undesirable. Additionally, the retrotransposons may be used to generate
XX      transgenic animals, to detect the presence of Candida in a sample, to
XX      detect and disrupt genes, and to assign functions to nucleotide
XX      sequences. The present sequence represents the C. albicans pcal
XX      retrotransposon.
XX
XX      Sequence 6426 BP; 2228 A; 1032 C; 1258 G; 1908 T; 0 other;
SQ

```

Query Match 100.0%; Score 6424.4; DB 21; Length 6426;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 6425; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

QY      1 tgttggttggtgactatttgtgtcagaactgataatgaatgagtattatga 60
QY      |||||||
DB      1 tgttggttggtgactatttgtgtcagaactgataatgaatgagtattatga 60
QY      61 gaatgaaaattttccatcacatcagtgatgacaaactaatctatttgtagt 120
QY      |||||||
DB      61 gaatgaaaattttccatcacatcagtgatgacaaactaatctatttgtagt 120
QY      121 ataataaggtatgaataccacaatccagaataatcaacgagatagaaggagagt 180
QY      |||||||
DB      121 ataataaggtatgaataccacaatccagaataatcaacgagatagaaggagagt 180
QY      181 tcaatatatactctgtgaataacttgcttctaattcaactaacacactagaagct 240
QY      |||||||

```


Db 181 tcaatataatctgtgtgaataaacttcgttcaatcctacatcacaaactagacgtgt 240
QY 241 acacgctcaatctcaggtlaaagaagtttatattcctacatcagatagagaagtcgtagat 300
Db 241 acacgctcaatctcaggtlaaagaagtttatattcctacatcagatagagaagtcgtagat 300
QY 301 aatcaattcgtcccaaatatagcggtgtataaatcaatccacgttcagatttgtatattgatt 360
Db 301 aatcaattcgtcccaaatatagcggtgtataaatcaatccacgttcagatttgtatattgatt 360
QY 361 gatagttcgaagtttgaaagttacagaaattccaaagatagattccgcaagaagtatga 420
Db 361 gatagttcgaagtttgaaagttacagaaattccaaagatagattccgcaagaagtatga 420
QY 421 taacgaagggaaggttcatggaaggtgtgtga tcaagctaaatgctatgtgaaggtgga 480
Db 421 taacgaagggaaggttcatggaaggtgtgtga tcaagctaaatgctatgtgaaggtgga 480
QY 481 acatataaagctcagatccaataatgcttccaataaaattatgacttacctaagtggc 540
Db 481 acatataaagctcagatccaataatgcttccaataaaattatgacttacctaagtggc 540
QY 541 cgtcgtgaatcagaagaagcgttgataaaatggaatggaatttaaatattccagcttgc 600
Db 541 cgtcgtgaatcagaagaagcgttgataaaatggaatggaatttaaatattccagcttgc 600
QY 601 ttaccccgatgttttgaaattttgtgctgaactaaatccaaagaataatccaaggttaa 660
Db 601 ttaccccgatgttttgaaattttgtgctgaactaaatccaaagaataatccaaggttaa 660
QY 661 aaaggtgaaggaatattatatttcttcgtgtgtgtgttlaacaaatggtttacagcttct 720
Db 661 aaaggtgaaggaatattatatttcttcgtgtgtgtgttlaacaaatggtttacagcttct 720
QY 721 tgaatgttcaagttgatactgaatttctaagctaccaaagcacttgcacaagaagcaaa 780
Db 721 tgaatgttcaagttgatactgaatttctaagctaccaaagcacttgcacaagaagcaaa 780
QY 781 cttaatacaagctcgtctatgatactgttactaaatccaaagattataccaattagtaa 840
Db 781 cttaatacaagctcgtctatgatactgttactaaatccaaagattataccaattagtaa 840
QY 841 gatctcgtcaagtttgataagttgaacatggaatgagtttggttataaacttcacaa 900
Db 841 gatctcgtcaagtttgataagttgaacatggaatgagtttggttataaacttcacaa 900
QY 901 ttgctgcaggttggaagaagaacttgaagaaataacttaacaacacttcaacgltgtga 960
Db 901 ttgctgcaggttggaagaagaacttgaagaaataacttaacaacacttcaacgltgtga 960
QY 961 tgaatgttgcaggttgaacttccaatctcattaggttcaagcttgcacttcaatcagtga 1020
Db 961 tgaatgttgcaggttgaacttccaatctcattaggttcaagcttgcacttcaatcagtga 1020
QY 1021 gaaatcagaagcttgaagttgttttgataatctacatgctcctactactcaaaagtggat 1080
Db 1021 gaaatcagaagcttgaagttgttttgataatctacatgctcctactactcaaaagtggat 1080
QY 1081 tcaagctgacaatgatacatcagtaactccaagtgtgcttaaccatagctgagaagaatgtg 1140
Db 1081 tcaagctgacaatgatacatcagtaactccaagtgtgcttaaccatagctgagaagaatgtg 1140
QY 1141 tgaatcctcgtatgatactgatactgtgtgatacttcaagaacaaataatgaaacttaact 1200
Db 1141 tgaatcctcgtatgatactgatactgtgtgatacttcaagaacaaataatgaaacttaact 1200
QY 1201 tatgtgtatgttaccagcaacagagaacaaagaagaagaacagagaagaaactcacttga 1260
Db 1201 tatgtgtatgttaccagcaacagagaacaaagaagaagaacagagaagaaactcacttga 1260
QY 1261 acaatctcaaaaagaagaacttgaatccaagaagaagaagaatgaagaacatccaatcaga 1320
Db 1261 acaatctcaaaaagaagaacttgaatccaagaagaagaagaatgaagaacatccaatcaga 1320

QY 1321 taacgaataagttgaaaaagaagaagaagaagaagaagaagaagaagaagaagaagaaga 1380
Db 1321 taacgaataagttgaaaaagaagaagaagaagaagaagaagaagaagaagaagaagaaga 1380
QY 1381 tgcgtctcattatgtgtatgataatcatatatttcgaacaaacagcttccagt 1440
Db 1381 tgcgtctcattatgtgtatgataatcatatatttcgaacaaacagcttccagt 1440
QY 1441 agaaaatctcattccttaattgcttcttgaaacgtatgaaattttaaaggttgaagtt 1500
Db 1441 agaaaatctcattccttaattgcttcttgaaacgtatgaaattttaaaggttgaagtt 1500
QY 1501 taacaagatcctgaagttgaatgatacaggtgcacaaatcctgttgcacaaataaagatat 1560
Db 1501 taacaagatcctgaagttgaatgatacaggtgcacaaatcctgttgcacaaataaagatat 1560
QY 1561 attgctgaatgttgaaggaacgaacaaatgaagtttctgttgcgtgaatgataatga 1620
Db 1561 attgctgaatgttgaaggaacgaacaaatgaagtttctgttgcgtgaatgataatga 1620
QY 1621 agcagatgtatgtgtatcctaaatcacaagagtcggtatgtctcagattacgttagaga 1680
Db 1621 agcagatgtatgtgtatcctaaatcacaagagtcggtatgtctcagattacgttagaga 1680
QY 1681 taacatgtatcctacagaagaagttccttaacccctgtgaggttgcacaatltgaagaacg 1740
Db 1681 taacatgtatcctacagaagaagttccttaacccctgtgaggttgcacaatltgaagaacg 1740
QY 1741 aggaattaaagtcttactatactaaagaatcagttgattttaaaccaaaatgtgtcc 1800
Db 1741 aggaattaaagtcttactatactaaagaatcagttgattttaaaccaaaatgtgtcc 1800
QY 1801 tactatattgtcttgaagaagaagaatgctgcgtgattcttatatggttccctcaatcagta 1860
Db 1801 tactatattgtcttgaagaagaagaatgctgcgtgattcttatatggttccctcaatcagta 1860
QY 1861 agaaatcttgaagaatgttgaatttgaatgaatgtgtgcacaaatgtgtccaaatgc 1920
Db 1861 agaaatcttgaagaatgttgaatttgaatgaatgtgtgcacaaatgtgtccaaatgc 1920
QY 1921 ccaagaatgaacaaagaataaacaagatgaatgaagaatgtcacaagaataccaagaatgata 1980
Db 1921 ccaagaatgaacaaagaataaacaagatgaatgaagaatgtcacaagaataccaagaatgata 1980
QY 1981 tagttcctgaagcattaaatcttcttgaacgaggttgaatttttaaggttgaatttc 2040
Db 1981 tagttcctgaagcattaaatcttcttgaacgaggttgaatttttaaggttgaatttc 2040
QY 2041 cccataatgaggttgaacaaatgtctacaaactgcagagataagaagatatttaattcca 2100
Db 2041 cccataatgaggttgaacaaatgtctacaaactgcagagataagaagatatttaattcca 2100
QY 2101 ttggtgtcaaatcatatgtcatttgaagaataatcttgttgaacaaaataccaaggtct 2160
Db 2101 ttggtgtcaaatcatatgtcatttgaagaataatcttgttgaacaaaataccaaggtct 2160
QY 2161 cgtactcaacacttcaaaaagaaggtctcacaagaatgtcgtatgtaaggtatgtcatt 2220
Db 2161 cgtactcaacacttcaaaaagaaggtctcacaagaatgtcgtatgtaaggtatgtcatt 2220
QY 2221 atcgaatgcacaacagagaagttcaacaatcatcatcagaagaagaagcctcgagaagaca 2280
Db 2221 atcgaatgcacaacagagaagttcaacaatcatcatcagaagaagaagcctcgagaagaca 2280
QY 2281 tgaagagcttcatgtgtatctcgtgcacattgaagtcgaagaataaagaatgtatlt 2340
Db 2281 tgaagagcttcatgtgtatctcgtgcacattgaagtcgaagaataaagaatgtatlt 2340
QY 2341 aaggtcgttatagatgatacaacaggttatactgaagaagaatttactaaagaacagaa 2400
Db 2341 aaggtcgttatagatgatacaacaggttatactgaagaagaatttactaaagaacagaa 2400

QY 2401 ggttaagagctcttaattcaacgataaagactctggaataatcggtttaacgataagt 2460
DB 2401 ggttaagagctcttaattcaacgataaagactctggaataatcggtttaacgataagt 2460
QY 2461 ggcatactcaagaagtataatgctctgagttcccaacacttcgatttaagtaagt 2520
DB 2461 ggcatactcaagaagtataatgctctgagttcccaacacttcgatttaagtaagt 2520
QY 2521 cggattctggaggagactatagcgacatctgacctgaagcttaagtctcgccgaagt 2580
DB 2521 cggattctggaggagactatagcgacatctgacctgaagcttaagtctcgccgaagt 2580
QY 2581 tgttaataaattgttttaacacagatttaacagatcggttgacacttggtccacaat 2640
DB 2581 tgttaataaattgttttaacacagatttaacagatcggttgacacttggtccacaat 2640
QY 2641 actaagttattattatgtatgtatcaattctcatcaatgatcaacacactccag 2700
DB 2641 actaagttattattatgtatgtatcaattctcatcaatgatcaacacactccag 2700
QY 2701 tcgttcaactcaaggagacaaacccctatggtgctatlaacatlaagtgagggaatt 2760
DB 2701 tcgttcaactcaaggagacaaacccctatggtgctatlaacatlaagtgagggaatt 2760
QY 2761 ctaccggtttccctttccatcgatgtgctgataattgaatgaatgcacgcaaaagaa 2820
DB 2761 ctaccggtttccctttccatcgatgtgctgataattgaatgaatgcacgcaaaagaa 2820
QY 2821 ccgttacagagttacatcaactaaaggagctccctcatcgatcgtggtgctgattg 2880
DB 2821 ccgttacagagttacatcaactaaaggagctccctcatcgatcgtggtgctgattg 2880
QY 2881 ctacgctagagattgctttagttatlaacggttgcttaaaaaatgctggtgataatt 2940
DB 2881 ctacgctagagattgctttagttatlaacggttgcttaaaaaatgctggtgataatt 2940
QY 2941 ccttaagccctaatgctgatatatgcgaagctagagttatlaaacctccatccaaana 3000
DB 2941 ccttaagccctaatgctgatatatgcgaagctagagttatlaaacctccatccaaana 3000
QY 3001 ctatccactacacatgatgatcacatgttctctatgctgaaggtalccagaggagca 3060
DB 3001 ctatccactacacatgatgatcacatgttctctatgctgaaggtalccagaggagca 3060
QY 3061 actgggagctcaatgaaggtacgagcaacatggtgaaagtgaataatgaatacaaaa 3120
DB 3061 actgggagctcaatgaaggtacgagcaacatggtgaaagtgaataatgaatacaaaa 3120
QY 3121 tgaagctgtagacatgcccagaagctatctagttcaagccagcatcggtttactaac 3180
DB 3121 tgaagctgtagacatgcccagaagctatctagttcaagccagcatcggtttactaac 3180
QY 3181 tacgggtacagttcttaacgataatgtataatgatgattccagttacagttacattga 3240
DB 3181 tacgggtacagttcttaacgataatgtataatgatgattccagttacagttacattga 3240
QY 3241 gaatccgagatcttcttaacactctcaactaactgaagaatccacagatagtatc 3300
DB 3241 gaatccgagatcttcttaacactctcaactaactgaagaatccacagatagtatc 3300
QY 3301 cgaagtaaatcggatgagatccctaaacccaggtcccaagcttaaaacccctgggataa 3360
DB 3301 cgaagtaaatcggatgagatccctaaacccaggtcccaagcttaaaacccctgggataa 3360
QY 3361 tccggtgctaaacctctcaactgtgacagagactcaataaggaagttaaga 3420
DB 3361 tccggtgctaaacctctcaactgtgacagagactcaataaggaagttaaga 3420
QY 3421 gacctatlaacaaacacaaagagagccctccatccagggggagagcaataagcct 3480
DB 3421 gacctatlaacaaacacaaagagagccctccatccagggggagagcaataagcct 3480
QY 3481 ggaactactgctcaggttgagatacaacaaacccagactggtactccgcttcgga 3540
DB 3481 ggaactactgctcaggttgagatacaacaaacccagactggtactccgcttcgga 3540
QY 3541 ggaactcaaatgttcaaggagacagatcattcggtgagcgtgtgtaagaacagcttc 3600
DB 3541 ggaactcaaatgttcaaggagacagatcattcggtgagcgtgtgtaagaacagcttc 3600
QY 3601 aagaattggcattctctgactaccggaactctgactgaggaatgagagcaaatcc 3660
DB 3601 aagaattggcattctctgactaccggaactctgactgaggaatgagagcaaatcc 3660
QY 3661 atctccgggggtgagtgatgactctgctgagcaactcaaatggttgagcaactaat 3720
DB 3661 atctccgggggtgagtgatgactctgctgagcaactcaaatggttgagcaactaat 3720
QY 3721 attccgggggtgagtgatgactctgctgagcaactcaaatggttgagcaactaat 3780
DB 3721 attccgggggtgagtgatgactctgctgagcaactcaaatggttgagcaactaat 3780
QY 3781 tgaacagagagtgctgtgtaacagtcacatccaaagaggttcccaaggaggaat 3840
DB 3781 tgaacagagagtgctgtgtaacagtcacatccaaagaggttcccaaggaggaat 3840
QY 3841 acttaagacaaactgatatagttgactggttctlaaagttatggaatgaaaaat 3900
DB 3841 acttaagacaaactgatatagttgactggttctlaaagttatggaatgaaaaat 3900
QY 3901 ctctccatlaattatgactgagatcactgagatcactgagatcactgagatcactgag 3960
DB 3901 ctctccatlaattatgactgagatcactgagatcactgagatcactgagatcactgag 3960
QY 3961 caattcaacagatccgaatcactcaatccaaatgaaatcaccaggtgataatgaa 4020
DB 3961 caattcaacagatccgaatcactcaatccaaatgaaatcaccaggtgataatgaa 4020
QY 4021 aaactgtaataatccaaacacatctgaagatctctgctgagtaagagattgagta 4080
DB 4021 aaactgtaataatccaaacacatctgaagatctctgctgagtaagagattgagta 4080
QY 4081 attggaacgatatagttgagaaattgagaaattgagaaattgagaaattgagaaatt 4140
DB 4081 attggaacgatatagttgagaaattgagaaattgagaaattgagaaattgagaaatt 4140
QY 4141 agctaatccactacacagatgaataatgataatgataatgataatgataatgataat 4200
DB 4141 agctaatccactacacagatgaataatgataatgataatgataatgataatgataat 4200
QY 4201 tcatagcatgltacagagcaagaagaalacacatttgaagaagaattgaacgaanaat 4260
DB 4201 tcatagcatgltacagagcaagaagaalacacatttgaagaagaattgaacgaanaat 4260
QY 4261 tgcgtgactaaacattcaactgatacactgatacactgatacactgatacactgatac 4320
DB 4261 tgcgtgactaaacattcaactgatacactgatacactgatacactgatacactgatac 4320
QY 4321 tactgtgaaacaaagagatcgaaacccaaagaaggggtgacctatcaactgtaaat 4380
DB 4321 tactgtgaaacaaagagatcgaaacccaaagaaggggtgacctatcaactgtaaat 4380
QY 4381 aaacaaagatcgaaataacagttacacatagttlaaagaagtgtgtagaaggttataa 4440
DB 4381 aaacaaagatcgaaataacagttacacatagttlaaagaagtgtgtagaaggttataa 4440
QY 4441 ccccaagaggtatttactacggtgltcaacaaatcgactaattgacccgggtgagat 4500
DB 4441 ccccaagaggtatttactacggtgltcaacaaatcgactaattgacccgggtgagat 4500
QY 4501 aaagtcaatgaatgctgaactgaagaatlttagatcaaaagtgtttagaagaagtcc 4560
DB 4501 aaagtcaatgaatgctgaactgaagaatlttagatcaaaagtgtttagaagaagtcc 4560
QY 4561 aattccacaggtgtgaactatctatggtttagtactactgaagaattgattc 4620
DB 4561 aattccacaggtgtgaactatctatggtttagtactactgaagaattgattc 4620

Db 4561 aattccaccggtgtgaagccatatactatgtgtgtgtacatactagaaaattgtctc 4620
 QY 4621 tctcaagggtgtgttcggaaatcacgtgtgtgtccatcggcacaacgaaaaggaaa 4680
 Db 4621 tctcaagggtgtgttcggaaatcacgtgtgtgtccatcggcacaacgaaaaggaaa 4680
 QY 4681 atgtgaatagaccctttagtgttgaatcacgtgtgtatagatctttagataaatt 4740
 Db 4681 atgtgaatagaccctttagtgttgaatcacgtgtgtatagatctttagataaatt 4740
 QY 4741 atgtgaatagaccctttagtgttgaatcacgtgtgtatagatctttagataaatt 4800
 Db 4741 atgtgaatagaccctttagtgttgaatcacgtgtgtatagatctttagataaatt 4800
 QY 4801 gtatcaaatgacctctatctcattcaaatccaattatgtcttctccctaactagt 4860
 Db 4801 gtatcaaatgacctctatctcattcaaatccaattatgtcttctccctaactagt 4860
 QY 4861 acccttgaagaaaaccatctgtgttattgaaaacgtctcgtctatgtgttaaacagtc 4920
 Db 4861 acccttgaagaaaaccatctgtgttattgaaaacgtctcgtctatgtgttaaacagtc 4920
 QY 4921 gggtgttgatgtgtatcacactatcaaaagatcttggaagacatgtgtttactcaagt 4980
 Db 4921 gggtgttgatgtgtatcacactatcaaaagatcttggaagacatgtgtttactcaagt 4980
 QY 4981 ttacacaatgatatgtttatttccatctgataatgaaaggatcggtatataattagg 5040
 Db 4981 ttacacaatgatatgtttatttccatctgataatgaaaggatcggtatataattagg 5040
 QY 5041 ttatatgttgatgtatcttctatgtgttggaagtcacaacaaagtattgataatttgc 5100
 Db 5041 ttatatgttgatgtatcttctatgtgttggaagtcacaacaaagtattgataatttgc 5100
 QY 5101 ggaatcattgagagatcatttgaagttaagtgtgttggaatataatcattctgtg 5160
 Db 5101 ggaatcattgagagatcatttgaagttaagtgtgttggaatataatcattctgtg 5160
 QY 5161 tattgaatttcgttaaaacggaatctgttataatttatacccaagaaaatttctcaagaa 5220
 Db 5161 tattgaatttcgttaaaacggaatctgttataatttatacccaagaaaatttctcaagaa 5220
 QY 5221 attactaaagatttcaaaactagatgactatatactggaataaacatccctgttcgaa 5280
 Db 5221 attactaaagatttcaaaactagatgactatatactggaataaacatccctgttcgaa 5280
 QY 5281 tgaacaatatgaaaaggttgcataatctcgtgaaaacgttaacccagaaatgatttga 5340
 Db 5281 tgaacaatatgaaaaggttgcataatctcgtgaaaacgttaacccagaaatgatttga 5340
 QY 5341 aaaggttcggaatgagacatctgttgaacctgtgtctaaaaaactatccaaagtgtgt 5400
 Db 5341 aaaggttcggaatgagacatctgttgaacctgtgtgtctaaaaaactatccaaagtgtgt 5400
 QY 5401 tggcgtgtcttatactggtgtgcacaaacacagctccagatalatcgtctgtagtaattc 5460
 Db 5401 tggcgtgtcttatactggtgtgcacaaacacagctccagatalatcgtctgtagtaattc 5460
 QY 5461 gttgggttctaaatctgcgaatccaaatgtccatgatttgaagaattgattatgtct 5520
 Db 5461 gttgggttctaaatctgcgaatccaaatgtccatgatttgaagaattgattatgtct 5520
 QY 5521 taggtatatacaaaatagatgagatatacatcttgagtaaaaagaagaagttgaatc 5580
 Db 5521 taggtatatacaaaatagatgagatatacatcttgagtaaaaagaagaagttgaatc 5580
 QY 5581 accacaaaatcatctgtatcgaatgttcagtgatgcgtatctgacccagatttga 5640
 Db 5581 accacaaaatcatctgtatcgaatgttcagtgatgcgtatctgacccagatttga 5640
 QY 5641 taagaataatcatagtggaactgtgatttgaatgaaattgtgtcaatggcgac 5700
 Db 5641 taagaataatcatagtggaactgtgatttgaatgaaattgtgtcaatggcgac 5700

QY 5701 caaaaaacaaacggtatagacaaagctcagacgtgtgtgaatgttggcttaatta 5760
 Db 5701 caaaaaacaaacggtatagacaaagctcagacgtgtgtgaatgttggcttaatta 5760
 QY 5761 tacaatgttgaagctatcgtgaataaaaaacattagattgtgtgttgaagtgg 5820
 Db 5761 tacaatgttgaagctatcgtgaataaaaaacattagattgtgtgttgaagtgg 5820
 QY 5821 taagatactgttcatcaagacacaaacgttgaattgaagtttgaataaactattg 5880
 Db 5821 taagatactgttcatcaagacacaaacgttgaattgaagtttgaataaactattg 5880
 QY 5881 tcaaccacaatcgacaatagatatcgtctataagtttctacgccaattgataatgaa 5940
 Db 5881 tcaaccacaatcgacaatagatatcgtctataagtttctacgccaattgataatgaa 5940
 QY 5941 agtatttcaatctcattatgttgaagcaaatgataattacgacctgttatgactaagt 6000
 Db 5941 agtatttcaatctcattatgttgaagcaaatgataattacgacctgttatgactaagt 6000
 QY 6001 tctaaagctgtcctaaatcacaagcattcgttgaaggtatgaaacggttggactaga 6060
 Db 6001 tctaaagctgtcctaaatcacaagcattcgttgaaggtatgaaacggttggactaga 6060
 QY 6061 agataatcaaacactgatacaaatgcaatacggcagaataagtgtatatactact 6120
 Db 6061 agataatcaaacactgatacaaatgcaatacggcagaataagtgtatatactact 6120
 QY 6121 attatcgtatgtctcaatcagaagggagttgtgtgtgtcactatttgtccaagaactg 6180
 Db 6121 attatcgtatgtctcaatcagaagggagttgtgtgtgtcactatttgtccaagaactg 6180
 QY 6181 atcaatgaaatgattgttattatgagaatggaataatttccatcacacatcagtgat 6240
 Db 6181 atcaatgaaatgattgttattatgagaatggaataatttccatcacacatcagtgat 6240
 QY 6241 gacgaacataaacctatttgtgtgtatgataaaggtatgataaataacacatcccaaga 6300
 Db 6241 gacgaacataaacctatttgtgtgtatgataaaggtatgataaataacacatcccaaga 6300
 QY 6301 tatcaacgagatagagaagggagtttcaatataatcttgtgaataaactcgttct 6360
 Db 6301 tatcaacgagatagagaagggagtttcaatataatcttgtgaataaactcgttct 6360
 QY 6361 aattactatatacaacactagacgtgttacaagctcaatctcaggttaagaagttatatt 6420
 Db 6361 aattactatatacaacactagacgtgttacaagctcaatctcaggttaagaagttatatt 6420
 QY 6421 ccatca 6426
 Db 6421 ccatca 6426

RESULT 2
 AA57936
 ID AA57936 standard; DNA; 392 BP.
 AC AA57936;
 XX
 DT 10-OCY-2000 (first entry)
 XX
 DE Candida albicans strain p36 TCa2 retrotransposon 5' region.
 XX
 KW Retrotransposon; pCal; TCa2; Ty1; copia; long terminal repeat; LTR;
 KW gag gene; group antigen; polyprotein; pol; aspartate protease; integrase;
 KW reverse transcriptase; RNaseH; pseudoknot; readthrough translation;
 KW stop codon suppression; gene delivery; gene therapy vector;
 KW genetic vaccine composition; immunogenic; transgenic animal; ds.
 OS Candida albicans strain p36.
 XX
 PN WO200026397-A1.

```

XX 11-MAY-2000.
PD
XX 01-NOV-1999;      99NC-NZ00179.
PR
XX 30-OCT-1998;      98CA-2249046.
PR
XX 30-OCT-1998;      98US-0106342.
PA      (JANC ) JANSSEN PHARM NV.
XX
XX Luyten WHML, De Backer MD, Nelissen BJM, Poultier RTM;
DR
XX WP1; 2000-365640/31.
PT
XX Novel retrotransposon expression vectors useful for expressing an
PT antigen, epitope or therapeutic agent, or detecting genes or the
PT presence of Candida in a sample -
XX
XX Example 9; Fig 9; 204pp; English.
XX
The invention relates to novel retrotransposons from the yeast Candida
alibicans which have a copy number of 40-150, preferably 50-100 copies
per genome. In particular, the invention relates to the novel C.
alibicans Ty1/copia retrotransposon pcal (AA57920), and to the integrated
form of this retrotransposon, designated TCa2, and to the novel C.
alibicans retrotransposons 1-28. pcal was initially isolated from C.
alibicans HOG1042 and has a copy number of 50-100 copies per cell. It
comprises identical 280 bp long terminal repeats (LTRs) and two open
reading frames (ORFs). The first ORF encodes a gag (group antigen)
protein, and the second ORF encodes a polyprotein (pol) consisting of an
aspartate protease, integrase, reverse transcriptase (RT) and RNaseH. The
gag and pol ORFs of pcal are in the same reading frame, separated only by
a termination codon (TGA). Translation of the pol ORF occurs through
the occasional readthrough suppression of the stop codon, which is
mediated by the formation of a pseudoknot within the gag-pol mRNA.
The retrotransposons of the invention can be used as vectors for in
vivo or in vitro transformation and expression. They can thus be used
for the delivery and expression of a therapeutic, immunological or
immunogenic molecule (e.g., an antigen) and may also be used for
eliciting an immunological response in a host organism. They are
therefore useful in genetic vaccine compositions and for gene therapy,
particularly where the use of retroviral vectors is unsafe or
undesirable. Additionally, the retrotransposons may be used to generate
transgenic animals, to detect the presence of candida in a sample, to
detect and disrupt genes, and to assign functions to nucleotide
sequences. Sequences AA57923-A57936 represents TCa2 retrotransposon 5'
regions from a variety of C. alibicans strains.
XX
Sequence 392 BP; 136 A; 56 C; 72 G; 128 T; 0 other:
XX
Query Match          6.1%; Score 392; DB 21; Length 392;
Best Local Similarity 100.0%; Pred. No. 1.5e-70;
Matches 392; Conservative 0; Mismatches 0; Indels 0; Gaps 0
QY   1 ttttggttgcacattttgttgttcgaacatcgatacagaatgaagtggattatga 60
DB   1 ttttggttgcactatttgcgtgcgaacgatcaaggaaagatggttatatga 60
QY   61 gaatgaaaattttccatcacacatcaagtgatgacaagaacaaactattgtagt 120
DB   61 gaatgaaaattttccatcacacatcaagtgatgacaagaacaaactattgtagt 120
QY   121 ataataagyggtatgaatatccaacatcccagaatatcaacagatagaaggagagt 180
DB   121 ataataagyggtatgaatatccaacatcccagaatatcaacagatagaaggagagt 180
QY   181 tcaatatatcttgytaataataacttcgtttaattcactatacaaacatgagcgigt 240
DB   181 tcaatatatcttgytaataataacttcgtttaattcactatacaaacatgagcgigt 240
QY   241 acacgtcaacttcaggttaaagaagtttatattccatccagattagaadtgcagtgat 300

```

Db	241	acagcgtccaaatcccgtaagaagatttatattccatccagattagagcgcgtagtgt	300
Oy	301	aatcaatctgcgtcccaaatatagcgtgtgtatataattcagtcctccaattgtattgatt	360
Db	301	aatcaatctgcgtcccaaatatagcgtgtgtatataattcagtcctccaattgtattgatt	360
Oy	361	gatagttcgaagtttgaagttacgaatttc	392
Db	361	gatagttcgaagtttgaagttacgaatttc	392
	RESULT	3	
ID	AAAS7923		
AC	AAAS7923	standard; DNA; 392 BP.	
XX	AAAS7923:		
DT	10-OCT-2000	(first entry)	
XX			
DE	Candida albicans strain SGV-1 TCA2 retrotransposon 5' region.		
XX			
KW	Retrotransposon: pCal; TCA2; TYL; copia; long terminal repeat; LTR;		
KW	gag gene; group antigen; polyprotein; pol; aspartate protease; integrase;		
KW	reverse transcriptase; RNaseH; pseudoknot; readthrough translation;		
KW	stop codon suppression; gene delivery; gene therapy vector;		
KW	genetic vaccine composition; immunogenic; transgenic animal; ds.		
XX			
OS	Candida albicans strain SGV-1.		
XX			
PN	WO200026397-A1.		
XX			
PD	11-MAY-2000.		
XX			
XX	01-NOV-1999;	99WO-NZ00179.	
PF			
XX			
PR	30-OCT-1998;	98CA-2249046.	
XX			
PR	30-OCT-1998;	98US-0106342.	
XX			
PA	(JANC) JANSSEN PHARM NV.		
PI	Luyten WHML, De Backer MD, Nelissen BJM, Poulter RTM;		
XX			
DR	WPI: 2000-365640/31.		
XX			
PT	Novel retrotransposon expression vectors useful for expressing an		
PT	antigen, epitope or therapeutic agent, or detecting genes or the		
PT	presence of Candida in a sample		
XX			
PS	Example 9; Fig 9; 204pp; English.		
XX			
CC	The invention relates to novel retrotransposons from the yeast Candida		
CC	albicans which have a copy number of 40-150, preferably 50-100 copies		
CC	per genome. In particular, the invention relates to the novel C.		
CC	albicans TY1/copia retrotransposon pCal (AA57920), and to the integrated		
CC	form of this retrotransposon, designated TCA2, and to the novel C.		
CC	albicans retrotransposons 1-28. pCal was initially isolated from C.		
CC	albicans HOG1042 and has a copy number of 50-100 copies per cell. It		
CC	comprises identical 280 bp long terminal repeats (LTRs) and two open		
CC	reading frames (ORFs). The first ORF encodes a gag (group antigen)		
CC	protein, and the second ORF encodes a polyprotein (pol) consisting of an		
CC	aspartate protease, integrase, reverse transcriptase (RT) and RNaseH. The		
CC	gag and pol ORFs of pCal are in the same reading frame, separated only by		
CC	a termination codon (TGA). Translation of the pol ORF occurs through		
CC	the occasional readthrough suppression of the stop codon, which is		
CC	mediated by the formation of a pseudoknot within the gag-pol mRNA.		
CC	The retrotransposons of the invention can be used as vectors for in		
CC	vivo or in vivo transformation and expression. They can thus be used		
CC	for the delivery and expression of a therapeutic, immunological or		
CC	immunogenic molecule (e.g., an antigen) and may also be used for		
CC	eliciting an immunological response in a host organism. They are		
CC	therefore useful in genetic vaccine compositions and for gene therapy,		
CC	particularly where the use of retroviral vectors is unsate or		
CC	undesirable. Additionally, the retrotransposons may be used to generate		

transgenic animals, to detect the presence of *Candida* in a sample, to detect and disrupt genes, and to assign functions to nucleotide sequences. Sequences AA57923-A57936 represents Tca2 retrotransposon 5' regions from a variety of *C. albicans* strains.

Sequence 392 BP; 136 A; 56 C; 72 G; 128 T; 0 other;

Query Match 6.1%; Score 388.8; DB 21; Length 392;
Best Local Similarity 99.5%; Pred. No. 6.9e-70;
Matches 390; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 tttgtgttgacactatttgggtgcagaactgcatacgaagaatgattattatga 60
DB 1 tttgtgttgacactatttgggtgcagaactgcatacgaagaatgattattatga 60
QY 61 gaatgaaattttccatcacacatcaggtgagatgacgaactaactattgtgtagt 120
DB 61 gaatgaaattttccatcacacatcaggtgagatgacgaactaactattgtgtagt 120
QY 121 ataaataagggtatgaataaccacaatccagaatatacagaagatgaaggagagt 180
DB 121 ataaataagggtatgaataaccacaatccagaatatacagaagatgaaggagagt 180
QY 181 tcaatatatcttctgtgaataataactcgttcttaattcactatcacaaactagacgtgt 240
DB 181 tcaatatatcttctgtgaataataactcgttcttaattcactatcacaaactagacgtgt 240
QY 241 aacagctcaatcctcaggtgaagaagaattatatacctcagattgagatgagatgatt 300
DB 241 aacagctcaatcctcaggtgaagaagaattatatacctcagattgagatgagatgatt 300
QY 301 aatcattcgtcccaaatgagcgttgtataaattcagtcctcagaattgtattatgatt 360
DB 301 aatcattcgtcccaaatgagcgttgtataaattcagtcctcagaattgtattatgatt 360
QY 361 gatagttcgaagtttgaaggtacagaatttc 392
DB 361 gatagttcgaagtttgaaggtacagaatttc 392

RESULT 4

AA57927
ID AA57927 standard; DNA: 392 BP.

AC AA57927;

DE 10-OCT-2000 (first entry)

XX *Candida albicans* strain ATC-1 Tca2 retrotransposon 5' region.

XX Retrotransposon; pCal; Tca2; Tyl; copia; long terminal repeat; LTR;
XX gag gene; group antigen; polyprotein; pol; aspartate protease; integrase;
XX reverse transcriptase; RNaseH; pseudoknot; readthrough translation;
XX stop codon suppression; gene delivery; gene therapy vector;
XX genetic vaccine composition; immunogenic; transgenic animal; ds.

XX *Candida albicans* strain ATC-1.

OS WO200026397-A1.

PN 11-MAY-2000.

PF 01-NOV-1999; 99WO-NZ00179.

PR 30-OCT-1998; 98CA-2249046.

PR 30-OCT-1998; 98US-0106342.

XX (JANNC) JANSSEN PHARM NV.

XX Luyten WHM, De Backer MD, Nelissen BJM, Poultier RTM;
XX WPI: 2000-365640/31.

XX Novel retrotransposon expression vectors useful for expressing an
PT antigen, epitope or therapeutic agent, or detecting genes or the
PT presence of *Candida* in a sample -

PS Example 9; Fig 9; 204pp; English.

XX The invention relates to novel retrotransposons from the yeast *Candida*
CC *albicans* which have a copy number of 40-150, preferably 50-100 copies
CC per genome. In particular, the invention relates to the novel C.
CC *albicans* Tyl/copia retrotransposon pCal (AA57920), and to the integrated
CC form of this retrotransposon, designated Tca2, and to the novel C.
CC *albicans* retrotransposons 1-28. pCal was initially isolated from C.
CC *albicans* H0G1042 and has a copy number of 50-100 copies per cell. It
CC comprises identical 280 bp long terminal repeats (LTRs) and two open
CC reading frames (ORFs). The first ORF encodes a gag (group antigen)
CC protein, and the second ORF encodes a polypeptide (pol) consisting of an
CC aspartate protease, integrase, reverse transcriptase (RT) and RNaseH. The
CC gag and pol ORFs of pCal are in the same reading frame, separated only by
CC a termination codon (TGA). Translation of the pol ORF occurs through
CC the occasional readthrough suppression of the stop codon, which is
CC mediated by the formation of a pseudoknot within the gag-pol mRNA.
CC The retrotransposons of the invention can be used as vectors for in
CC vitro or in vivo transformation and expression. They can thus be used
CC for the delivery and expression of a therapeutic, immunological or
CC immunogenic molecule (e.g., an antigen) and may also be used for
CC eliciting an immunological response in a host organism. They are
CC therefore useful in genetic vaccine compositions and for gene therapy,
CC particularly where the use of retroviral vectors is unsafe or
CC undesirable. Additionally, the retrotransposons may be used to generate
CC transgenic animals, to detect the presence of *Candida* in a sample, to
CC detect and disrupt genes, and to assign functions to nucleotide
CC sequences. Sequences AA57923-A57936 represents Tca2 retrotransposon 5'
CC regions from a variety of *C. albicans* strains.

XX Sequence 392 BP; 136 A; 56 C; 72 G; 128 T; 0 other;

Query Match 6.1%; Score 388.8; DB 21; Length 392;
Best Local Similarity 99.5%; Pred. No. 6.9e-70;
Matches 390; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 tttgtgttgacactatttgggtgcagaactgcatacgaagaatgattattatga 60
DB 1 tttgtgttgacactatttgggtgcagaactgcatacgaagaatgattattatga 60
QY 61 gaatgaaattttccatcacacatcaggtgagatgacgaactaactattgtgtagt 120
DB 61 gaatgaaattttccatcacacatcaggtgagatgacgaactaactattgtgtagt 120
QY 121 ataaataagggtatgaataaccacaatccagaatatacagaagatgaaggagagt 180
DB 121 ataaataagggtatgaataaccacaatccagaatatacagaagatgaaggagagt 180
QY 181 tcaatatatcttctgtgaataataactcgttcttaattcactatcacaaactagacgtgt 240
DB 181 tcaatatatcttctgtgaataataactcgttcttaattcactatcacaaactagacgtgt 240
QY 241 aacagctcaatcctcaggtgaagaagaattatatacctcagattgagatgagatgatt 300
DB 241 aacagctcaatcctcaggtgaagaagaattatatacctcagattgagatgagatgatt 300
QY 301 aatcattcgtcccaaatgagcgttgtataaattcagtcctcagaattgtattatgatt 360
DB 301 aatcattcgtcccaaatgagcgttgtataaattcagtcctcagaattgtattatgatt 360
QY 361 gatagttcgaagtttgaaggtacagaatttc 392
DB 361 gatagttcgaagtttgaaggtacagaatttc 392

RESULT 5

AA57928

ID AAA57928 standard; DNA: 392 BP.
 AC AAA57928;
 DT 10-OCT-2000 (first entry)
 DE Candida albicans strain ATC-2 TCa2 retrotransposon 5' region.
 KW Retrotransposon; pCal; TCa2; TY1; copia; long terminal repeat; LTR;
 KW gag gene; group antigen; polyprotein; pol; aspartate protease; integrase;
 KW reverse transcriptase; RNaseH; pseudoknot; readthrough translation;
 KW stop codon suppression; gene delivery; gene therapy vector;
 KW genetic vaccine composition; immunogenic; transgenic animal; ds.
 OS Candida albicans strain ATC-2.
 PN WO200026397-A1.
 PD 11-MAY-2000.
 XX 01-NOV-1999; 99WO-NZ00179.
 XX 30-OCT-1998; 98CA-2249046.
 XX 30-OCT-1998; 98US-0106342.
 XX (JANCS) JANSSEN PHARM NV.
 XX Luyten WHML, De Backer MD, Nelissen BJM, Poultier RTM;
 XX WPI: 2000-365640/31.
 PT Novel retrotransposon expression vectors useful for expressing an
 PT antigen, epitope or therapeutic agent, or detecting genes or the
 PT presence of Candida in a sample -
 PS Example 9; Fig 9; 204pp; English.
 XX The invention relates to novel retrotransposons from the yeast Candida
 CC albicans which have a copy number of 40-150, preferably 50-100 copies
 CC per genome. In particular, the invention relates to the novel C.
 CC albicans TY1/copia retrotransposon pCal (AAA57920), and to the integrated
 CC form of this retrotransposon, designated TCa2, and to the novel C.
 CC albicans retrotransposons 1-28. pCal was initially isolated from C.
 CC albicans h031042 and has a copy number of 50-100 copies per cell. It
 CC comprises identical 280 bp long terminal repeats (LTRs) and two open
 CC reading frames (ORFs). The first ORF encodes a gag (group antigen)
 CC protein, and the second ORF encodes a polyprotein (pol) consisting of an
 CC aspartate protease, integrase, reverse transcriptase (RT) and RNaseH. The
 CC gag and pol ORFs of pCal are in the same reading frame, separated only by
 CC a termination codon (TGA). Translation of the pol ORF occurs through
 CC the occasional readthrough suppression of the stop codon, which is
 CC mediated by the formation of a pseudoknot within the gag-pol mRNA.
 CC The retrotransposons of the invention can be used as vectors for in
 CC vitro or in vivo transformation and expression. They can thus be used
 CC for the delivery and expression of a therapeutic, immunological or
 CC immunogenic molecule (e.g., an antigen) and may also be used for
 CC eliciting an immunological response in a host organism. They are
 CC therefore useful in genetic vaccine compositions and for gene therapy,
 CC particularly where the use of retroviral vectors is unsafe or
 CC undesirable. Additionally, the retrotransposons may be used to generate
 CC transgenic animals, to detect the presence of Candida in a sample, to
 CC detect and disrupt genes, and to assign functions to nucleotide
 CC sequences. Sequences AAA57923-A57936 represents TCa2 retrotransposon 5'
 CC regions from a variety of C. albicans strains.
 SQ Sequence 392 BP; 136 A; 56 C; 72 G; 128 T; 0 other;

Query Match 6.1%; Score 388.8; DB 21; Length 392;
 Best Local Similarity 99.5%; Pred. No. 6.9e-70;
 Matches 390; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 tgggtgttgacactatttgggtcgaagaactgacatgaatgaatgattattatga 60

Db 1 tgggtgttgacactatttgggtcgaagaactgacatgaatgaatgattattatga 60
 QY 61 gaatgaaatattttccatcacacatcagtgatgacgaactaaactatattgtgt 120
 Db 61 gaatgaaatattttccatcacacatcagtgatgacgaactaaactatattgtgt 120
 QY 121 ataaataaggatgaataaccacacatccagaataccaagatagaaggagagt 180
 Db 121 ataaataaggatgaataaccacacatccagaataccaagatagaaggagagt 180
 QY 181 tcaatatactcttgaataaacttgcgttaacttaactatacacaaactagagtg 240
 Db 181 tcaatatactcttgaataaacttgcgttaacttaactatacacaaactagagtg 240
 QY 241 acagcctcaatctcagtgaaagaatttatatccatagattagaagtgatgt 300
 Db 241 acagcctcaatctcagtgaaagaatttatatccatagattagaagtgatgt 300
 QY 301 aatcattcgtcccaaataggcgttgcataaattcagctccagaatttcatattgatt 360
 Db 301 aatcattcgtcccaaataggcgttgcataaattcagctccagaatttcatattgatt 360
 QY 361 gatagttcgaagtttgaaggtacagaatttc 392
 Db 361 gatagttcgaagtttgaaggtacagaatttc 392

RESULT 6

ID AAA57931 standard; DNA: 392 BP.
 AC AAA57931;
 DT 10-OCT-2000 (first entry)
 DE Candida albicans strain F16-1 TCa2 retrotransposon 5' region.
 KW Retrotransposon; pCal; TCa2; TY1; copia; long terminal repeat; LTR;
 KW gag gene; group antigen; polyprotein; pol; aspartate protease; integrase;
 KW reverse transcriptase; RNaseH; pseudoknot; readthrough translation;
 KW stop codon suppression; gene delivery; gene therapy vector;
 KW genetic vaccine composition; immunogenic; transgenic animal; ds.
 OS Candida albicans strain F16-1.
 PN WO200026397-A1.
 PD 11-MAY-2000.
 XX 01-NOV-1999; 99WO-NZ00179.
 XX 30-OCT-1998; 98CA-2249046.
 XX 30-OCT-1998; 98US-0106342.
 XX (JANCS) JANSSEN PHARM NV.
 XX Luyten WHML, De Backer MD, Nelissen BJM, Poultier RTM;
 XX WPI: 2000-365640/31.
 PT Novel retrotransposon expression vectors useful for expressing an
 PT antigen, epitope or therapeutic agent, or detecting genes or the
 PT presence of Candida in a sample -
 PS Example 9; Fig 9; 204pp; English.
 XX The invention relates to novel retrotransposons from the yeast Candida
 CC albicans which have a copy number of 40-150, preferably 50-100 copies
 CC per genome. In particular, the invention relates to the novel C.
 CC albicans TY1/copia retrotransposon pCal (AAA57920), and to the integrated
 CC form of this retrotransposon, designated TCa2, and to the novel C.
 CC albicans retrotransposons 1-28. pCal was initially isolated from C.

CC albicans h061042 and has a copy number of 50-100 copies per cell. It
 CC comprises identical 280 bp long terminal repeats (LTRs) and two open
 CC reading frames (ORFs). The first ORF encodes a gag (group antigen)
 CC protein, and the second ORF encodes a polypeptide (pol) consisting of an
 CC aspartate protease, integrase, reverse transcriptase (RT) and RNaseH. The
 CC gag and pol ORFs of pcal are in the same reading frame, separated only by
 CC a termination codon (TGA). Translation of the pol ORF occurs through
 CC the occasional readthrough suppression of the stop codon, which is
 CC mediated by the formation of a pseudoknot within the gag-pol mRNA.
 CC The retrotransposons of the invention can be used as vectors for in
 CC vitro or in vivo transformation and expression. They can thus be used
 CC for the delivery and expression of a therapeutic, immunological or
 CC immunogenic molecule (e.g., an antigen) and may also be used for
 CC eliciting an immunological response in a host organism. They are
 CC therefore useful in genetic vaccine compositions and for gene therapy,
 CC particularly where the use of retroviral vectors is unsafe or
 CC undesirable. Additionally, the retrotransposons may be used to generate
 CC transgenic animals, to detect the presence of Candida in a sample, to
 CC detect and disrupt genes, and to assign functions to nucleotide
 CC sequences. Sequences AA57923-AA57936 represents TCA2 retrotransposon 5'
 CC regions from a variety of C. albicans strains.

SQ Sequence 392 BP; 136 A; 56 C; 72 G; 128 T; 0 other;

Query Match 6.1%; Score 388.8; DB 21; Length 392;
 Best Local Similarity 99.5%; Pred. No. 6.9e-70;
 Matches 390; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 tttgtgttgcactatttgcgcgaacatgataagaagatggtattatga 60
 DB 1 tttgtgttgcactatttgcgcgaacatgataagaagatggtattatga 60
 QY 61 gaatgaaattttccatcacatcagtgatgacagactaatatattgtatg 120
 DB 61 gaatgaaattttccatcacatcagtgatgacagactaatatattgtatg 120
 QY 121 ataataaggatgaataccacatccagatatacaagagatagaaggagagt 180
 DB 121 ataataaggatgaataccacatccagatatacaagagatagaaggagagt 180
 QY 121 ataataaggatgaataccacatccagatatacaagagatagaaggagagt 180
 DB 121 ataataaggatgaataccacatccagatatacaagagatagaaggagagt 180
 QY 181 tcaatatacttctgtgaataaactcgttcaatcactatacaactagaagtt 240
 DB 181 tcaatatacttctgtgaataaactcgttcaatcactatacaactagaagtt 240
 QY 241 acacgctcaatcgcagtaagaagtttatattcattcagattagaagtcagat 300
 DB 241 acacgctcaatcgcagtaagaagtttatattcattcagattagaagtcagat 300
 QY 301 aatcaatttgcctccaatttagcgtttataaattcagtcctcagatttattatgatt 360
 DB 301 aatcaatttgcctccaatttagcgtttataaattcagtcctcagatttattatgatt 360
 QY 361 gatagtttgcgaatttgaagttacgaatttc 392
 DB 361 gatagtttgcgaatttgaagttacgaatttc 392

RESULT 7
 ID AA57932 standard: DNA; 392 BP.
 AC AA57932;
 DE 10-OCT-2000 (first entry)
 XX Candida albicans strain Fl6-2 TCA2 retrotransposon 5' region.
 KW Retrotransposon: pCal; TCA2; TY1; copia; long terminal repeat; LTR;
 KW gag gene; group antigen; polypeptide; pol; aspartate protease; integrase;
 KW reverse transcriptase; RNaseH; pseudoknot; readthrough translation;
 KW stop codon suppression; gene delivery; gene therapy vector;
 KW genetic vaccine composition; immunogenic; transgenic animal; ds.

XX OS Candida albicans strain Fl6-2.
 XX PN WO200026397-A1.
 XX PD 11-MAY-2000.
 XX PF 01-NOV-1999; 99WO-NZ00179.
 XX PR 30-OCT-1998; 98CA-2249046.
 XX PR 30-OCT-1998; 98US-0106342.
 XX PA (JANC) JANSSEN PHARM NV.
 XX PI Luyten WHM, De Backer MD, Nelissen BJM, Poultier RTM.
 XX DR WPI; 2000-365640/31.
 XX PT Novel retrotransposon expression vectors useful for expressing an
 XX antigen, epitope or therapeutic agent, or detecting genes or the
 XX presence of candida in a sample
 PS Example 9; Fig 9; 204pp; English.

CC The invention relates to novel retrotransposons from the yeast Candida
 CC albicans which have a copy number of 40-150, preferably 50-100 copies
 CC per genome. In particular the invention relates to the novel C.
 CC albicans TY1/copia retrotransposon pcal (AA57920), and to the integrated
 CC form of this retrotransposon, designated TCA2, and to the novel C.
 CC albicans retrotransposons 1-28. pcal was initially isolated from C.
 CC albicans h061042 and has a copy number of 50-100 copies per cell. It
 CC comprises identical 280 bp long terminal repeats (LTRs) and two open
 CC reading frames (ORFs). The first ORF encodes a gag (group antigen)
 CC protein, and the second ORF encodes a polypeptide (pol) consisting of an
 CC aspartate protease, integrase, reverse transcriptase (RT) and RNaseH. The
 CC gag and pol ORFs of pcal are in the same reading frame, separated only by
 CC a termination codon (TGA). Translation of the pol ORF occurs through
 CC the occasional readthrough suppression of the stop codon, which is
 CC mediated by the formation of a pseudoknot within the gag-pol mRNA.
 CC The retrotransposons of the invention can be used as vectors for in
 CC vitro or in vivo transformation and expression. They can thus be used
 CC for the delivery and expression of a therapeutic, immunological or
 CC immunogenic molecule (e.g., an antigen) and may also be used for
 CC eliciting an immunological response in a host organism. They are
 CC therefore useful in genetic vaccine compositions and for gene therapy,
 CC particularly where the use of retroviral vectors is unsafe or
 CC undesirable. Additionally, the retrotransposons may be used to generate
 CC transgenic animals, to detect the presence of Candida in a sample, to
 CC detect and disrupt genes, and to assign functions to nucleotide
 CC sequences. Sequences AA57923-AA57936 represents TCA2 retrotransposon 5'
 CC regions from a variety of C. albicans strains.

SQ Sequence 392 BP; 136 A; 56 C; 72 G; 128 T; 0 other;

Query Match 6.1%; Score 388.8; DB 21; Length 392;
 Best Local Similarity 99.5%; Pred. No. 6.9e-70;
 Matches 390; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 tttgtgttgcactatttgcgcgaacatgataagaagatggtattatga 60
 DB 1 tttgtgttgcactatttgcgcgaacatgataagaagatggtattatga 60
 QY 61 gaatgaaattttccatcacatcagtgatgacagactaatatattgtatg 120
 DB 61 gaatgaaattttccatcacatcagtgatgacagactaatatattgtatg 120
 QY 121 ataataaggatgaataccacatccagatatacaagagatagaaggagagt 180
 DB 121 ataataaggatgaataccacatccagatatacaagagatagaaggagagt 180
 QY 181 tcaatatacttctgtgaataaactcgttcaatcactatacaactagaagtt 240
 DB 181 tcaatatacttctgtgaataaactcgttcaatcactatacaactagaagtt 240

Db 181 tcaatataatctgtgataataacttcgttcaatccatcacacaactagacgt 240
 Qy 241 aacgcctcaatctcagtgaaagaagtttatccatccatgagatgcagatgac 300
 Db 241 aacgcctcaatctcagtgaaagaagtttatccatccatgagatgcagatgac 300
 Qy 301 aatcatttcgtcccaaatcagctgtgataaatccatccatgagatgcagatgac 360
 Db 301 aatcatttcgtcccaaatcagctgtgataaatccatccatgagatgcagatgac 360
 Qy 361 gatagttcgaagttgaagttacagaatttc 392
 Db 361 gatagttcgaagttgaagttacagaatttc 392
 RESULT 8
 AAAS7925
 ID AAAS7925 standard; DNA: 392 BP.
 AC AAAS7925;
 XX 10-OCT-2000 (first entry)
 DE Candida albicans strain SC5-1 Tca2 retrotransposon 5' region.
 XX Retrotransposon: pCal; Tca2; Ty1; copia; long terminal repeat; LTR;
 KW gag gene; group antigen; polyprotein; pol; aspartate protease; integrase;
 KW reverse transcriptase; RNaseH; pseudoknot; readthrough translation;
 KW stop codon suppression; gene delivery; gene therapy vector;
 KW genetic vaccine composition; immunogenic; transgenic animal; ds.
 XX Candida albicans strain SC5-1.
 OS WO200026397-A1.
 PN 11-MAY-2000.
 PD 01-NOV-1999; 99MO-NZ00179.
 PE 30-OCT-1998; 98CA-2249046.
 PR 30-OCT-1998; 98US-0106342.
 XX (JANC) JANSSEN PHARM NV.
 PA Luyten WHML, De Backer MD, Nelissen BJM, Poulter RTM.
 PI WPI: 2000-365640/31.
 DR Novel retrotransposon expression vectors useful for expressing an
 PT antigen, epitope or therapeutic agent, or detecting genes or the
 PS presence of Candida in a sample -
 PS Example 9; Fig 9; 204pp; English.
 XX The invention relates to novel retrotransposons from the yeast Candida
 CC albicans which have a copy number of 40-150, preferably 50-100 copies
 CC per genome. In particular, the invention relates to the novel C.
 CC albicans Ty1/copia retrotransposon pCal (AAAS7920), and to the integrated
 CC form of this retrotransposon, designated Tca2, and to the novel C.
 CC albicans retrotransposons 1-28. pCal was initially isolated from C.
 CC albicans HOG1042 and has a copy number of 50-100 copies per cell. It
 CC comprises identical 280 bp long terminal repeats (LTRs) and two open
 CC reading frames (ORFs). The first ORF encodes a gag (group antigen)
 CC protein, and the second ORF encodes a polyprotein (pol) consisting of an
 CC aspartate protease, integrase, reverse transcriptase (RT) and RNaseH. The
 CC gag and pol ORFs of pCal are in the same reading frame, separated only by
 CC a termination codon (TGA). Translation of the pol ORF occurs through
 CC the occasional readthrough suppression of the stop codon, which is
 CC mediated by the formation of a pseudoknot within the gag-pol mRNA.
 CC The retrotransposons of the invention can be used as vectors for in
 CC vitro or in vivo transformation and expression. They can thus be used
 CC for the delivery and expression of a therapeutic, immunological or
 CC immunogenic molecule (e.g., an antigen) and may also be used for

CC eliciting an immunological response in a host organism. They are
 CC therefore useful in genetic vaccine compositions and for gene therapy.
 CC particularly where the use of retroviral vectors is unsate or
 CC undesirable. Additionally, the retrotransposons may be used to generate
 CC transgenic animals, to detect the presence of Candida in a sample, to
 CC detect and disrupt genes, and to assign functions to nucleotide
 CC sequences. Sequences AAAS7923-A57936 represents Tca2 retrotransposon 5'
 CC regions from a variety of C. albicans strains.
 XX Sequence 392 BP; 136 A; 56 C; 71 G; 129 T; 0 other;
 SQ
 Query Match 6.0%; Score 387.2; DB 21; Length 392;
 Best Local Similarity 99.2%; Pred. No. 1.5e-69;
 Matches 389; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 Qy 1 tttgtgttgacactatttgcgcagaactgcatgaatgaatgattgtattatga 60
 Db 1 tttgtgttgacactatttgcgcagaactgcatgaatgaatgattgtattatga 60
 Qy 61 gaatggaatatttccatcacacatcagtgatgacagaactaacatattgtgtagt 120
 Db 61 gaatggaatatttccatcacacatcagtgatgacagaactaacatattgtgtagt 120
 Qy 121 ataataaaggatgaataaccacacatccagaatatacaagagatagaagagagatt 180
 Db 121 ataataaaggatgaataaccacacatccagaatatacaagagatagaagagagatt 180
 Qy 181 tcaatataatctgtgataataacttcgttcaatccatcacacaactagacgtgt 240
 Db 181 tcaatataatctgtgataataacttcgttcaatccatcacacaactagacgtgt 240
 Qy 241 aacgcctcaatctcagtgaaagaagtttatccatccatgagatgcagatgac 300
 Db 241 aacgcctcaatctcagtgaaagaagtttatccatccatgagatgcagatgac 300
 Qy 301 aatcatttcgtcccaaatcagctgtgataaatccatccatgagatgcagatgac 360
 Db 301 aatcatttcgtcccaaatcagctgtgataaatccatccatgagatgcagatgac 360
 Qy 361 gatagttcgaagttgaagttacagaatttc 392
 Db 361 gatagttcgaagttgaagttacagaatttc 392
 RESULT 9
 AAAS7929
 ID AAAS7929 standard; DNA: 392 BP.
 AC AAAS7929;
 XX 10-OCT-2000 (first entry)
 DE Candida albicans strain SA4-1 Tca2 retrotransposon 5' region.
 XX Retrotransposon: pCal; Tca2; Ty1; copia; long terminal repeat; LTR;
 KW gag gene; group antigen; polyprotein; pol; aspartate protease; integrase;
 KW reverse transcriptase; RNaseH; pseudoknot; readthrough translation;
 KW stop codon suppression; gene delivery; gene therapy vector;
 KW genetic vaccine composition; immunogenic; transgenic animal; ds.
 XX Candida albicans strain SA4-1.
 OS WO200026397-A1.
 PN 11-MAY-2000.
 PD 01-NOV-1999; 99MO-NZ00179.
 PE 30-OCT-1998; 98CA-2249046.
 PR 30-OCT-1998; 98US-0106342.
 XX (JANC) JANSSEN PHARM NV.

XX Luyten WHML, De Backer MD, Nelissen BJM, Poulter RTM;
 XX WPI; 2000-365640/31.
 XX
 PT Novel retrotransposon expression vectors useful for expressing an
 PT antigen, epitope or therapeutic agent, or detecting genes or the
 PT presence of *Candida* in a sample -
 PS
 PS Example 9; Fig 9; 204pp; English.
 XX
 CC The invention relates to novel retrotransposons from the yeast *Candida*
 CC albicans which have a copy number of 40-150, preferably 50-100 copies
 CC per genome. In particular, the invention relates to the novel C.
 CC albicans Tyl1/copia retrotransposon pcal (AA57920), and to the integrated
 CC form of this retrotransposon, designated Tca2, and to the novel C.
 CC albicans retrotransposons 1-28. pcal was initially isolated from C.
 CC albicans HOG1042 and has a copy number of 50-100 copies per cell. It
 CC comprises identical 280 bp long terminal repeats (LTRs) and two open
 CC reading frames (ORFs). The first ORF encodes a gag (group antigen)
 CC protein, and the second ORF encodes a polypeptide (pol) consisting of an
 CC aspartate protease, integrase, reverse transcriptase (RT) and RNaseH. The
 CC gag and pol ORFs of pcal are in the same reading frame, separated only by
 CC a termination codon (TGA). Translation of the pol ORF occurs through
 CC the occasional readthrough suppression of the stop codon, which is
 CC mediated by the formation of a pseudoknot within the gag-pol mRNA.
 CC The retrotransposons of the invention can be used as vectors for in
 CC vitro or in vivo transformation and expression. They can thus be used
 CC for the delivery and expression of a therapeutic, immunological or
 CC immunogenic molecule (e.g., an antigen) and may also be used for
 CC eliciting an immunological response in a host organism. They are
 CC therefore useful in genetic vaccine compositions and for gene therapy,
 CC particularly where the use of retroviral vectors is unsafe or
 CC undesirable. Additionally, the retrotransposons may be used to generate
 CC transgenic animals, to detect the presence of *Candida* in a sample, to
 CC detect and disrupt genes, and to assign functions to nucleotide
 CC sequences. Sequences AA57923-AA57936 represents Tca2 retrotransposon 5'
 CC regions from a variety of C. albicans strains.
 XX
 XX Sequence 392 BP; 135 A; 57 C; 72 G; 128 T; 0 other;
 SO
 Query Match 6.0%; Score 387.2; DB 21; Length 392;
 Best Local Similarity 99.2%; Pred. No. 1.5e-69;
 Matches 389; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

RESULT 10
 ID AA57933
 XX AA57933 standard; DNA; 392 BP.
 AC
 AC AA57933;
 XX
 XX 10-OCT-2000 (first entry)
 DE
 DE *Candida albicans* strain 759-1 Tca2 retrotransposon 5' region.
 XX
 XX Retrotransposon; pcal; Tca2; Tyl1; copia; long terminal repeat; LTR;
 KW gag gene; group antigen; polyprotein; pol; aspartate protease; integrase;
 KW reverse transcriptase; RNaseH; pseudoknot; readthrough translation;
 KW stop codon suppression; gene delivery; gene therapy vector;
 KW genetic vaccine composition; immunogenic; transgenic animal; ds.
 XX
 XX *Candida albicans* strain 759-1.
 OS
 XX MO200026397-A1.
 XX
 XX 11-MAY-2000.
 PD
 XX 01-NOV-1999; 99WO-N200179.
 XX
 XX 30-OCT-1998; 98CA-2249046.
 XX
 XX 30-OCT-1998; 98US-0106342.
 PR
 XX (JAN) JANSSEN PHARM NV.
 PA
 PA Luyten WHML, De Backer MD, Nelissen BJM, Poulter RTM;
 PI WPI; 2000-365640/31.
 XX
 XX Novel retrotransposon expression vectors useful for expressing an
 PT antigen, epitope or therapeutic agent, or detecting genes or the
 PT presence of *Candida* in a sample -
 PT
 PT Example 9; Fig 9; 204pp; English.
 PS
 XX The invention relates to novel retrotransposons from the yeast *Candida*
 CC albicans which have a copy number of 40-150, preferably 50-100 copies
 CC per genome. In particular, the invention relates to the novel C.
 CC albicans Tyl1/copia retrotransposon pcal (AA57920), and to the integrated
 CC form of this retrotransposon, designated Tca2, and to the novel C.
 CC albicans retrotransposons 1-28. pcal was initially isolated from C.
 CC albicans HOG1042 and has a copy number of 50-100 copies per cell. It
 CC comprises identical 280 bp long terminal repeats (LTRs) and two open
 CC reading frames (ORFs). The first ORF encodes a gag (group antigen)
 CC protein, and the second ORF encodes a polypeptide (pol) consisting of an
 CC aspartate protease, integrase, reverse transcriptase (RT) and RNaseH. The
 CC gag and pol ORFs of pcal are in the same reading frame, separated only by
 CC a termination codon (TGA). Translation of the pol ORF occurs through
 CC the occasional readthrough suppression of the stop codon, which is
 CC mediated by the formation of a pseudoknot within the gag-pol mRNA.
 CC The retrotransposons of the invention can be used as vectors for in
 CC vitro or in vivo transformation and expression. They can thus be used
 CC for the delivery and expression of a therapeutic, immunological or
 CC immunogenic molecule (e.g., an antigen) and may also be used for
 CC eliciting an immunological response in a host organism. They are
 CC therefore useful in genetic vaccine compositions and for gene therapy,
 CC particularly where the use of retroviral vectors is unsafe or
 CC undesirable. Additionally, the retrotransposons may be used to generate
 CC transgenic animals, to detect the presence of *Candida* in a sample, to
 CC detect and disrupt genes, and to assign functions to nucleotide
 CC sequences. Sequences AA57923-AA57936 represents Tca2 retrotransposon 5'
 CC regions from a variety of C. albicans strains.
 CC
 XX Sequence 392 BP; 138 A; 57 C; 71 G; 126 T; 0 other;
 SO

Query Match 6.0%; Score 387.2; DB 21; Length 392;

Best Local Similarity 99.2%; Pred. No. 1.5e-69; Matches 389; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 tttgtgttgcacattttgttcagaactatcaatgaagaatgtattatga 60
 |||||||
 Db 1 tttgtgttgcacattttgttcagaactatcaatgaagaatgtattatga 60
 QY 61 gaatgaaatatttccatcacacatcaggtgtagacagaactaataatattgtgtagt 120
 |||||||
 Db 61 gaatgaaatatttccatcacacatcaggtgtagacagaactaataatattgtgtagt 120
 QY 121 ataataagggtatgaataatccacatccagaatatacgaagatagaaggagagatt 180
 |||||||
 Db 121 ataataagggtatgaataatccacatccagaatatacgaagatagaaggagagatt 180
 QY 181 tcaatataatctgtgaataatcgtttcaattcactatacacaactagacgtgt 240
 |||||||
 Db 181 tcaatataatctgtgtgaataatcgtttcaattcactatacacaactagacgtgt 240
 QY 241 acacgctcaatctcaggtaaagaagatttatccatcacatcagattagaagtcagatgtat 300
 |||||||
 Db 241 acacgctcaatctcaggtaaagaagatttatccatcacatcagattagaagtcagatgtat 300
 QY 301 aataattcgtcccaaataggcgtgtataaattcagtcctcagattgtattatgtatt 360
 |||||||
 Db 301 aataattcgtcccaaataggcgtgtataaattcagtcctcagattgtattatgtatt 360
 QY 361 gatagttcgaagttgaagtcagaatttc 392
 |||||||
 Db 361 gatagttcgaagttgaagtcagaatttc 392

RESULT 11

AAAS7930
 ID AAAS7930 standard; DNA; 392 BP.

AC AAAS7930;

DT 10-OCT-2000 (first entry)

DE Candida albicans strain SA4-2 Tca2 retrotransposon 5' region.

XX Retrotransposon; pCal; Tca2; Tyl; copia; long terminal repeat; LTR;
 KW gag gene; group antigen; polyprotein; pol; aspartate protease; integrase;
 KW reverse transcriptase; RNaseH; pseudoknot; readthrough translation;
 KW stop codon suppression; gene delivery; gene therapy vector;
 KW genetic vaccine composition; immunogenic; transgenic animal; ds.

XX Candida albicans strain SA4-2.

CS WO200026397-A1.

PD 11-MAY-2000.

XX 01-NOV-1999; 99WO-NZ00179.

XX 30-OCT-1998; 98CA-2249046.

XX 30-OCT-1998; 98US-0106342.

XX (JAMC) JANSSEN PHARM NV.

XX Luyten WHML, De Backer MD, Nelissen BJM, Foulter RTM;

XX WPI; 2000-365640/31.

XX Novel retrotransposon expression vectors useful for expressing an

XX antigen, epitope or therapeutic agent, or detecting genes or the

XX presence of Candida in a sample -

XX Example 9; Fig 9; 204pp; English.

XX The invention relates to novel retrotransposons from the yeast Candida

XX albicans which have a copy number of 40-150, preferably 50-100 copies

CC per genome. In particular, the invention relates to the novel C.
 CC albicans Tyl1/copia retrotransposon pCal (AAAS7920), and to the integrated
 CC form of this retrotransposon, designated Tca2, and to the novel C.
 CC albicans retrotransposons 1-28. pCal was initially isolated from C.
 CC albicans HOG1042 and has a copy number of 50-100 copies per cell. It
 CC comprises identical 280 bp long terminal repeats (LTRs) and two open
 CC reading frames (ORFs). The first ORF encodes a gag (group antigen)
 CC protein, and the second ORF encodes a polyprotein (pol) consisting of an
 CC aspartate protease, integrase, reverse transcriptase (RT) and RNaseH. The
 CC gag and pol ORFs of pCal are in the same reading frame, separated only by
 CC a termination codon (TGA). Translation of the pol ORF occurs through
 CC the occasional readthrough suppression of the stop codon, which is
 CC mediated by the formation of a pseudoknot within the gag-pol mRNA.
 CC The retrotransposons of the invention can be used as vectors for in
 CC vitro or in vivo transformation and expression. They can thus be used
 CC for the delivery and expression of a therapeutic, immunological or
 CC immunogenic molecule (e.g., an antigen) and may also be used for
 CC eliciting an immunological response in a host organism. They are
 CC therefore useful in genetic vaccine compositions and for gene therapy,
 CC particularly where the use of retroviral vectors is unsafe or
 CC undesirable. Additionally, the retrotransposons may be used to generate
 CC transgenic animals, to detect the presence of Candida in a sample, to
 CC detect and disrupt genes, and to assign functions to nucleotide
 CC sequences. Sequences AAAS7923-AA57936 represents Tca2 retrotransposon 5'
 CC regions from a variety of C. albicans strains.

XX Sequence 392 BP; 135 A; 58 C; 72 G; 127 T; 0 other;

Query Match 6.0%; Score 385.6; DB 21; Length 392;

Best Local Similarity 99.0%; Pred. No. 3.1e-69; Matches 388; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 tttgtgttgcacattttgttcagaactgatacgaagaatgtattatga 60
 |||||||
 Db 1 tttgtgttgcacattttgttcagaactgatacgaagaatgtattatga 60
 QY 61 gaatgaaatatttccatcacacatcaggtgtagacagaactaataatattgtgtagt 120
 |||||||
 Db 61 gaatgaaatatttccatcacacatcaggtgtagacagaactaataatattgtgtagt 120
 QY 121 ataataagggtatgaataatccacatccagaatatacgaagatagaaggagagatt 180
 |||||||
 Db 121 ataataagggtatgaataatccacatccagaatatacgaagatagaaggagagatt 180
 QY 181 tcaatataatctgtgaataatcgtttcaattcactatacacaactagacgtgt 240
 |||||||
 Db 181 tcaatataatctgtgtgaataatcgtttcaattcactatacacaactagacgtgt 240
 QY 241 acacgctcaatctcaggtaaagaagatttatccatcacatcagattagaagtcagatgtat 300
 |||||||
 Db 241 acacgctcaatctcaggtaaagaagatttatccatcacatcagattagaagtcagatgtat 300
 QY 301 aataattcgtcccaaataggcgtgtataaattcagtcctcagattgtattatgtatt 360
 |||||||
 Db 301 aataattcgtcccaaataggcgtgtataaattcagtcctcagattgtattatgtatt 360
 QY 361 gatagttcgaagttgaagtcagaatttc 392
 |||||||
 Db 361 gatagttcgaagttgaagtcagaatttc 392

RESULT 12

AAAS7934
 ID AAAS7934 standard; DNA; 392 BP.

AC AAAS7934;

DT 10-OCT-2000 (first entry)

DE Candida albicans strain 759-2 Tca2 retrotransposon 5' region.

XX Retrotransposon; pCal; Tca2; Tyl; copia; long terminal repeat; LTR;

gag gen. group antigen; pol: aspartate protease; integrase;
 KM reverse transcriptase; RnaseH; pseudonoc; readthrough translation;
 KM stop codon suppression; gene delivery; gene therapy vector;
 KM genetic vaccine composition; immunogenic; transgenic animal; ds.
 OS Candida albicans strain 759-2.
 PM WO200026397-A1.
 PD 11-MAY-2000.
 PF 01-NOV-1999; 99WO-N200179.
 PR 30-OCT-1998; 98CA-2249046.
 PR 30-OCT-1998; 98US-0106342.
 XX (JANC) JANSSEN PHARM NV.
 XX Luyten WHML, De Backer MD, Nelissen BJW, Poultier RTM;
 DR WPI: 2000-365640/31.
 XX Novel retrovirusposon expression vectors useful for expressing an
 PT antigen, epitope or therapeutic agent, or detecting genes or the
 PT presence of Candida in a sample -
 PS Example 9; Fig 9; 204pp; English.
 XX The invention relates to novel retrovirusposons from the yeast Candida
 CC albicans which have a copy number of 40-150, preferably 50-100 copies
 CC per genome. In particular, the invention relates to the novel C.
 CC albicans Ty1/copia retrovirusposon pcal (AA57920), and to the integrated
 CC form of this retrovirusposon, designated TCa2, and to the novel C.
 CC albicans retrovirusposons 1-28. pcal was initially isolated from C.
 CC albicans hoi1042 and has a copy number of 50-100 copies per cell. It
 CC comprises identical 280 bp long terminal repeats (LTRs) and two open
 CC reading frames (ORFs). The first ORF encodes a gag (group antigen)
 CC protein, and the second ORF encodes a polypeptide (pol) consisting of an
 CC aspartate protease, integrase, reverse transcriptase (RT) and RnaseH. The
 CC gag and pol ORFs of pcal are in the same reading frame, separated only by
 CC a termination codon (TGA). Translation of the pol ORF occurs through
 CC the occasional readthrough suppression of the stop codon, which is
 CC mediated by the formation of a pseudoknot within the gag-pol mRNA.
 CC The retrovirusposons of the invention can be used as vectors for in
 CC vitro or in vivo transformation and expression. They can thus be used
 CC for the delivery and expression of a therapeutic, immunological or
 CC immunogenic molecule (e.g., an antigen) and may also be used for
 CC eliciting an immunological response in a host organism. They are
 CC therefore useful in genetic vaccine compositions and for gene therapy,
 CC particularly where the use of retroviral vectors is unsafe or
 CC undesirable. Additionally, the retrovirusposons may be used to generate
 CC transgenic animals, to detect the presence of Candida in a sample, to
 CC detect and disrupt genes, and to assign functions to nucleotide
 CC sequences. Sequences AA57923-57936 represents TCa2 retrovirusposon 5'
 CC regions from a variety of C. albicans strains.
 XX

Db	121	atcaataagggatgaaataaccacatccccaagatatacaacgagatagaaggaggagact	180
Qy	181	tcaatatatactctgtgtaataactctgttctaattcaactatacacaactagacgtgt	240
Db	181	tcaatatatactctgtgtaataactctgttctaattcaactatacacaactagacgtgt	240
Qy	241	acaagctcaactcgaaggaagaagtttatccatcagattagaaatcgatagtt	300
Db	241	acaagctcaactcgaaggaagaagtttatccatcagattagaaatcgatagtt	300
Qy	301	aatcaatctgcgccaaattagtcgtgtatataatcagtcctcagattgtatattgatt	360
Db	301	aacccattcgcgccaaattagtcgtgtatataatcagtcctcagattgtatattgatt	360
Qy	361	gatagttcgaagttgaaagtaagaatttc	392
Db	361	gatagttcgaagttgaaagtaagaatttc	392
RESULT 13			
AAAS7935	ID	AAAS7935 standard; DNA; 392 BP.	
XX	AAAS7935:		
AC	10-OCT-2000 (first entry)		
XX			
DE	Candida albicans strain p30 TCa2 retrotransposon 5' region.		
XX			
KM	Retrotransposon: pcal; TCa2; Tyl; copia; long terminal repeat; LTR;		
KM	gag gene; group antigen; polyprotein; pol; aspartate protease; integrase;		
KM	reverse transcriptase; RNaseH; pseudoknot; readthrough translation;		
KM	stop codon suppression; gene delivery; gene therapy vector;		
KM	genetic vaccine composition; immunogenic; transgenic animal; ds.		
XX			
OS	Candida albicans strain p30.		
XX			
PM	W0200026397-A1.		
PD	11-MAY-2000.		
XX			
PF	01-NOV-1999; 99WC-NZ00179.		
XX			
PR	30-OCT-1998; 98CA-2249046.		
PR	30-OCT-1998; 98US-0106342.		
XX			
PA	(JANC) JANSSEN PHARM NV.		
PI	Luyten WHML, De Backer MD, Nelissen BJM, Poulter RTM;		
XX			
XX	WPI: 2000-365640/31.		
XX			
PT	Novel retrotransposon expression vectors useful for expressing an		
PT	antigen, epitope or therapeutic agent, or detecting genes or the		
XX	presence of Candida in a sample -		
XX			
PS	Example 9; Fig 9; 204pp; English.		
XX			
CC	The invention relates to novel retrotransposons from the yeast Candida		
CC	albicans which have a copy number of 40-150, preferably 50-100 copies		
CC	per genome. In particular, the invention relates to the novel C.		
CC	albicans Tyl/copia retrotransposon pcal (AAAS7920), and to the integrated		
CC	form of this retrotransposon, designated TCa2, and to the novel C.		
CC	albicans retrotransposons 1-28. pcal was initially isolated from C.		
CC	albicans HOG1042 and has a copy number of 50-100 copies per cell. It		
CC	comprises identical 280 bp long terminal repeats (LTRs) and two open		
CC	reading frames (ORFs). The first ORF encodes a gag (group antigen)		
CC	protein, and the second ORF encodes a polyprotein (pol) consisting of an		
CC	aspartate protease, integrase, reverse transcriptase (RT) and RNaseH. The		
CC	gag and pol ORFs of pcal are in the same reading frame, separated only by		
CC	a termination codon (TGA). Translation of the pol ORF occurs through		
CC	the occasional readthrough suppression of the stop codon, which is		
CC	mediated by the formation of a pseudoknot within the gag-pol mRNA.		

CC The retrotransposons of the invention can be used as vectors for in
CC vitro or in vivo transformation and expression. They can thus be used
CC for the delivery and expression of a therapeutic, immunological or
CC immunogenic molecule (e.g., an antigen) and may also be used for
CC eliciting an immunological response in a host organism. They are
CC therefore useful in genetic vaccine compositions and for gene therapy,
CC particularly where the use of retroviral vectors is unsafe or
CC undesirable. Additionally, the retrotransposons may be used to generate
CC transgenic animals, to detect the presence of *Candida* in a sample, to
CC detect and disrupt genes, and to assign functions to nucleotide
CC sequences. Sequences AA57923-AS7936 represents TCA2 retrotransposon 5'
CC regions from a variety of *C. albicans* strains.

XX Sequence 392 BP; 139 A; 55 C; 71 G; 127 T; 0 other;

Query Match 6.0%; Score 384; DB 21; Length 392;
Best Local Similarity 98.7%; Pred. No. 6,5e-69;
Matches 387; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

```

QY 1 tttgtgttgcacattttgtgtcagaactgataatgaatgatgtattatga 60
   1 tttgtgttgcacattttgtgtcagaactgataatgaatgatgtattatga 60
   1 tttgtgttgcacattttgtgtcagaactgataatgaatgatgtattatga 60
DB 61 gaatgaaatatttccatcacacatcagtgatgacagaactaactatattgtagt 120
   61 gaatgaaatatttccatcacacatcagtgatgacagaactaactatattgtagt 120
QY 121 ataataagggtatgaataaccacatcccgatataccaacgagatagaaggagaggt 180
   121 ataataagggtatgaataaccacatcccgatataccaacgagatagaaggagaggt 180
DB 121 ataataagggtatgaataaccacatcccgatataccaacgagatagaaggagaggt 180
   121 ataataagggtatgaataaccacatcccgatataccaacgagatagaaggagaggt 180
QY 181 tcaatataatctgtgataataactcgttcaattcactaactaaccacatagcgt 240
   181 tcaatataatctgtgataataactcgttcaattcactaactaaccacatagcgt 240
DB 181 tcaatataatctgtgataataactcgttcaattcactaactaaccacatagcgt 240
   181 tcaatataatctgtgataataactcgttcaattcactaactaaccacatagcgt 240
QY 241 acacgctcaatcagtgaaagaagtttatattccatccagatagaagtcagatgat 300
   241 acacgctcaatcagtgaaagaagtttatattccatccagatagaagtcagatgat 300
DB 241 acacgctcaatcagtgaaagaagtttatattccatccagatagaagtcagatgat 300
   241 acacgctcaatcagtgaaagaagtttatattccatccagatagaagtcagatgat 300
QY 301 aatcattcgtcccaaatcagtgatataatccagtcctcagatattgtattgatt 360
   301 aatcattcgtcccaaatcagtgatataatccagtcctcagatattgtattgatt 360
DB 301 aatcattcgtcccaaatcagtgatataatccagtcctcagatattgtattgatt 360
   301 aatcattcgtcccaaatcagtgatataatccagtcctcagatattgtattgatt 360
QY 361 gatagttcgaagttgaagtagaatttc 392
   361 gatagttcgaagttgaagtagaatttc 392
DB 361 gatagttcgaagttgaagtagaatttc 392
   361 gatagttcgaagttgaagtagaatttc 392

```

SUIT 14

AA57924 standard; DNA; 392 BP.

AA57924;

10-OCT-2000 (first entry)

Candida albicans strain SGV-2 TCA2 retrotransposon 5' region.

Retroransposon; pcal; TCA2; TY1; copia; long terminal repeat; LTR;
gag gene; group antigen; polyprotein; pol; aspartate protease; integrase;
reverse transcriptase; RnaseH; pseudoknot; readthrough translation;
stop codon suppression; gene delivery; gene therapy vector;
genetic vaccine composition; immunogenic; transgenic animal; ds.

Candida albicans strain SGV-2.

WO200026397-A1.

11-MAY-2000.

01-NOV-1999; 99WO-NZ00179.

XX

PR 30-OCT-1998; 98CA-2249046.

PR 30-OCT-1998; 98US-0106342.

XX (JANC) JANSSEN PHARM NV.

XX Luyten WHML, De Backer MD, Nelissen BJM, Poulier RTM;

XX WPI; 2000-365640/31.

PT Novel retrotransposon expression vectors useful for expressing an
PT antigen, epitope or therapeutic agent, or detecting genes or the
PT presence of *Candida* in a sample

XX Example 9; Fig 9; 204pp; English.

CC The invention relates to novel retrotransposons from the yeast *Candida*
CC *albicans* which have a copy number of 40-150, preferably 50-100 copies
CC per genome. In particular, the invention relates to the novel *C.*
CC *albicans* TY1/copia retrotransposon pcal (AA57920), and to the integrated
CC form of this retrotransposon, designated TCA2, and to the novel *C.*
CC *albicans* retrotransposons 1-28. pcal was initially isolated from *C.*
CC *albicans* HOG1042, and has a copy number of 50-100 copies per cell. It
CC comprises identical 280 bp long terminal repeats (LTRs) and two open
CC reading frames (ORFs). The first ORF encodes a gag (group antigen)
CC protein, and the second ORF encodes a polyprotein (pol) consisting of an
CC aspartate protease, integrase, reverse transcriptase (RT) and RnaseH. The
CC gag and pol ORFs of pcal are in the same reading frame, separated only by
CC a termination codon (TGA). Translation of the pol ORF occurs through
CC the occasional readthrough suppression of the stop codon, which is
CC mediated by the formation of a pseudoknot within the gag-pol mRNA.
CC The retrotransposons of the invention can be used as vectors for in
CC vitro or in vivo transformation and expression. They can thus be used
CC for the delivery and expression of a therapeutic, immunological or
CC immunogenic molecule (e.g., an antigen) and may also be used for
CC eliciting an immunological response in a host organism. They are
CC therefore useful in genetic vaccine compositions and for gene therapy,
CC particularly where the use of retroviral vectors is unsafe or
CC undesirable. Additionally, the retrotransposons may be used to generate
CC transgenic animals, to detect the presence of *Candida* in a sample, to
CC detect and disrupt genes, and to assign functions to nucleotide
CC sequences. Sequences AA57923-AS7936 represents TCA2 retrotransposon 5'
CC regions from a variety of *C. albicans* strains.

SQ Sequence 392 BP; 135 A; 57 C; 74 G; 126 T; 0 other;

Query Match 6.0%; Score 382.4; DB 21; Length 392;
Best Local Similarity 98.5%; Pred. No. 1.4e-68;
Matches 386; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

```

QY 1 tttgtgttgcacattttgtgtcagaactgataatgaatgatgtattatga 60
   1 tttgtgttgcacattttgtgtcagaactgataatgaatgatgtattatga 60
   1 tttgtgttgcacattttgtgtcagaactgataatgaatgatgtattatga 60
DB 61 gaatgaaatatttccatcacacatcagtgatgacagaactaactatattgtagt 120
   61 gaatgaaatatttccatcacacatcagtgatgacagaactaactatattgtagt 120
QY 61 gaatgaaatatttccatcacacatcagtgatgacagaactaactatattgtagt 120
   61 gaatgaaatatttccatcacacatcagtgatgacagaactaactatattgtagt 120
DB 61 gaatgaaatatttccatcacacatcagtgatgacagaactaactatattgtagt 120
   61 gaatgaaatatttccatcacacatcagtgatgacagaactaactatattgtagt 120
QY 121 ataataagggtatgaataaccacatcccgatataccaacgagatagaaggagaggt 180
   121 ataataagggtatgaataaccacatcccgatataccaacgagatagaaggagaggt 180
DB 121 ataataagggtatgaataaccacatcccgatataccaacgagatagaaggagaggt 180
   121 ataataagggtatgaataaccacatcccgatataccaacgagatagaaggagaggt 180
QY 181 tcaatataatctgtgataataactcgttcaattcactaactaaccacatagcgt 240
   181 tcaatataatctgtgataataactcgttcaattcactaactaaccacatagcgt 240
DB 181 tcaatataatctgtgataataactcgttcaattcactaactaaccacatagcgt 240
   181 tcaatataatctgtgataataactcgttcaattcactaactaaccacatagcgt 240
QY 241 acacgctcaatcagtgaaagaagtttatattccatccagatagaagtcagatgat 300
   241 acacgctcaatcagtgaaagaagtttatattccatccagatagaagtcagatgat 300
DB 241 acacgctcaatcagtgaaagaagtttatattccatccagatagaagtcagatgat 300
   241 acacgctcaatcagtgaaagaagtttatattccatccagatagaagtcagatgat 300
QY 301 aatcattcgtcccaaatcagtgatataatccagtcctcagatattgtattgatt 360
   301 aatcattcgtcccaaatcagtgatataatccagtcctcagatattgtattgatt 360
DB 301 aatcattcgtcccaaatcagtgatataatccagtcctcagatattgtattgatt 360
   301 aatcattcgtcccaaatcagtgatataatccagtcctcagatattgtattgatt 360

```

QY 361 gatagttcgaagttgaagtcacaagaatttc 392
 ||||||| ||||||| ||||||| |||||||
 Db 361 gatagttcgaagttgaagtcacaagaatttc 392

RESULT 15

AAAS7926
 ID AAAS7926 standard; DNA; 392 BP.

XX AAAS7926;
 AC

XX 10-OCT-2000 (first entry)
 DT

DE Candida albicans strain SC5-2 Tca2 retrotransposon 5' region.

KW Retrotransposon; pCal; Tca2; TY1; copia; long terminal repeat; LTR;
 KW gag gene; group antigen; polyprotein; pol; aspartate protease; integrase;
 KW reverse transcriptase; RNaseH; pseudoknot; readthrough translation;
 KW stop codon suppression; gene delivery; gene therapy vector;
 KW genetic vaccine composition; immunogenic; transgenic animal; ds.

OS Candida albicans strain SC5-2.

PN WO200026397-A1.

PD 11-MAY-2000.

PE 01-NOV-1999; 99WO-NZ00179.

PR 30-OCT-1998; 98CA-2249046.

PR 30-OCT-1998; 98US-0106342.

PA (JANC) JANSSEN PHARM NV.

PI Luyten WHML, De Backer MD, Nelissen BJM, Poulter RTM;

DR WPI; 2000-365640/31.

PT Novel retrotransposon expression vectors useful for expressing an
 PT antigen, epitope or therapeutic agent, or detecting genes or the
 PT presence of Candida in a sample -

PS Example 9; Fig 9; 204pp; English.

XX The invention relates to novel retrotransposons from the yeast Candida
 CC albicans which have a copy number of 40-150, preferably 50-100 copies
 CC per genome. In particular, the invention relates to the novel C.
 CC albicans TY1/copia retrotransposon pCal (AAAS7920), and to the integrated
 CC form of this retrotransposon, designated Tca2, and to the novel C.
 CC albicans retrotransposons 1-28. pCal was initially isolated from C.
 CC albicans hog1042 and has a copy number of 50-100 copies per cell. It
 CC comprises identical 280 bp long terminal repeats (LTRs) and two open
 CC reading frames (ORFs). The first ORF encodes a gag (group antigen)
 CC protein, and the second ORF encodes a polyprotein (pol) consisting of an
 CC aspartate protease, integrase, reverse transcriptase (RT) and RNaseH. The
 CC gag and pol ORFs of pCal are in the same reading frame, separated only by
 CC a termination codon (TGA). Translation of the pol ORF occurs through
 CC the occasional readthrough suppression of the stop codon, which is
 CC mediated by the formation of a pseudoknot within the gag-pol mRNA.
 CC The retrotransposons of the invention can be used as vectors for in
 CC vitro or in vivo transformation and expression. They can thus be used
 CC for the delivery and expression of a therapeutic, immunological or
 CC immunogenic molecule (e.g., an antigen) and may also be used for
 CC eliciting an immunological response in a host organism. They are
 CC therefore useful in genetic vaccine compositions and for gene therapy,
 CC particularly where the use of retroviral vectors is unsafe or
 CC undesirable. Additionally, the retrotransposons may be used to generate
 CC transgenic animals, to detect the presence of Candida in a sample, to
 CC detect and disrupt genes, and to assign functions to nucleotide
 CC sequences. Sequences AAAS7923-A57936 represents Tca2 retrotransposon 5'
 CC regions from a variety of C. albicans strains.

SQ Sequence 392 BP; 137 A; 57 C; 71 G; 127 T; 0 other;

Query Match 5.9%; Score 380.8; DB 21; Length 392;

Best Local Similarity 96.2%; Pred. No. 2.9e-66;

Matches 385; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 tcttggttgcacactatttgcagaaactgacatgaataatgatgttatatga 60
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
 Db 1 tcttggttgcacactatttgcagaaactgacatgaataatgatgttatatga 60
 QY 61 gaatggaataatttccatccacacatcagtgatgacagaactaactatattgtagt 120
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
 Db 61 gaatggaataatttccatccacacatcagtgatgacagaactaactatattgtagt 120
 QY 121 ataataagggtatgaataatcaccacatcccgaaatcacaagatagaaggagaggtt 180
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
 Db 121 ataataagggtatgaataatcaccacatcccgaaatcacaagatagaaggagaggtt 180
 QY 181 tcaatataatcttgatataataacttcgttcaattcactatcacacaactagaagctgt 240
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
 Db 181 tcaatataatcttgatataataacttcgttcaattcactatcacacaactagaagctgt 240
 QY 241 acaagctcaatcctcaggtaaagaaagtttatatccatccagattagaagtcgatgtgat 300
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
 Db 241 acaagctcaatcctcaggtaaagaaagtttatatccatccagattagaagtcgaagcgtat 300
 QY 301 aatcatttcgtcccaaatatagcgttgttaaatccagccccaatttgttatgtatt 360
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
 Db 301 aatcatttcgtcccaaatatagcgttgttaaatccagccccaatttgttatgtatt 360
 QY 361 gatagttcgaagtttgaagtcacaagaatttc 392
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
 Db 361 gatagttcgaagtttgaagtcacaagaatttc 392

Search completed: August 25, 2001, 02:35:04
 Job time: 13259 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 24, 2001, 22:20:55 ; Search time 4490.13 seconds
(without alignments)
13528.347 Million cell updates/sec

Title: US-09-430-590E-3
Perfect score: 6426
Sequence: 1 tgttggttggtgcacattt.....agaagttatattccatca 6426

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 10228115 seqs, 4726426750 residues 20456230
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*

1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est33:*
26: gb_est34:*
27: gb_est35:*
28: gb_est36:*
29: gb_est37:*
30: gb_est38:*
31: gb_est39:*
32: gb_est40:*
33: em_estba:*
34: em_estfun:*
35: em_esthum1:*
36: em_esthum2:*
37: em_esthum3:*
38: em_esthum4:*
39: em_esthum5:*
40: em_esthum6:*
41: em_esthum7:*
42: em_esthum8:*
43: em_esthum9:*

44: em_esthum10:*
45: em_esthum11:*
46: em_esthum12:*
47: em_esthum13:*
48: em_esthum14:*
49: em_esthum15:*
50: em_esthum16:*
51: em_esthum17:*
52: em_esthum18:*
53: em_esthum19:*
54: em_esthum20:*
55: em_esthum21:*
56: em_esthum22:*
57: em_esthum23:*
58: em_esthum24:*
59: em_esthum25:*
60: em_esthum26:*
61: em_esthum27:*
62: em_esthum28:*
63: em_estin1:*
64: em_estin2:*
65: em_estin3:*
66: em_estin4:*
67: em_estin5:*
68: em_estom1:*
69: em_estov1:*
70: em_estov2:*
71: em_estpl1:*
72: em_estpl2:*
73: em_estpl3:*
74: em_estpl4:*
75: em_estpl5:*
76: em_estpl6:*
77: em_estpl7:*
78: em_estpl8:*
79: em_estpl9:*
80: em_estpl10:*
81: em_estro1:*
82: em_estro2:*
83: em_estro3:*
84: em_estro4:*
85: em_estro5:*
86: em_estro6:*
87: em_estro7:*
88: em_estro8:*
89: em_estro9:*
90: em_estro10:*
91: em_estro11:*
92: em_estro12:*
93: em_estro13:*
94: em_estro14:*
95: em_estro15:*
96: em_estro16:*
97: em_estro17:*
98: em_estro18:*
99: em_estro19:*
100: em_estro20:*
101: gb_est25:*
102: gb_est26:*
103: gb_est27:*
104: gb_est28:*
105: gb_est29:*
106: gb_est30:*
107: gb_est31:*
108: gb_est32:*
109: gb_est41:*
110: gb_est42:*
111: gb_est43:*
112: gb_est44:*
113: gb_est45:*
114: gb_est46:*
115: gb_est47:*
116:

117: gb_est48:*
 118: gb_est49:*
 119: gb_est50:*
 120: gb_est51:*
 121: gb_est52:*
 122: gb_est53:*
 123: gb_est54:*
 124: gb_est55:*
 125: gb_est56:*
 126: gb_est57:*
 127: gb_est58:*
 128: gb_est59:*
 129: gb_est60:*
 130: gb_est61:*
 131: gb_est62:*
 132: gb_est63:*
 133: gb_est64:*
 134: gb_est65:*
 135: gb_est66:*
 136: gb_est67:*
 137: gb_est68:*
 138: gb_est69:*
 139: gb_est70:*
 140: gb_est71:*
 141: gb_est72:*
 142: gb_est73:*
 143: gb_est74:*
 144: gb_est75:*
 145: gb_est76:*
 146: gb_est77:*
 147: gb_est78:*
 148: gb_est79:*
 149: gb_est80:*
 150: gb_est81:*
 151: gb_est82:*
 152: gb_est83:*
 153: gb_est84:*
 154: gb_est85:*
 155: gb_est86:*
 156: gb_est87:*
 157: gb_est88:*
 158: gb_est89:*
 159: gb_est90:*
 160: gb_est91:*
 161: gb_est92:*
 162: gb_est93:*
 163: gb_est94:*
 164: gb_est95:*
 165: gb_est96:*
 166: gb_est97:*
 167: gb_est98:*
 168: gb_est99:*
 169: gb_est100:*
 170: gb_est101:*
 171: gb_est102:*
 172: gb_est103:*
 173: gb_est104:*
 174: gb_est105:*
 175: gb_est106:*
 176: gb_est107:*
 177: gb_est108:*
 178: gb_est109:*
 179: gb_est110:*
 180: gb_est111:*
 181: gb_est112:*
 182: gb_est113:*
 183: gb_est114:*
 184: gb_est115:*
 185: gb_est116:*
 186: gb_est117:*
 187: gb_est118:*
 188: gb_est119:*
 189: gb_est120:*

190: gb_est110:*
 191: gb_est111:*
 192: gb_est112:*
 193: gb_est113:*
 194: gb_est114:*
 195: gb_est115:*
 196: gb_est116:*
 197: gb_est117:*
 198: gb_est118:*
 199: gb_est119:*
 200: gb_est120:*
 201: gb_est121:*
 202: gb_est122:*
 203: gb_est123:*
 204: gb_est124:*
 205: gb_est125:*
 206: gb_est126:*
 207: gb_est127:*
 208: gb_est128:*
 209: gb_est129:*
 210: gb_est130:*
 211: gb_est131:*
 212: gb_est132:*
 213: gb_est133:*
 214: gb_est134:*
 215: gb_est135:*
 216: gb_est136:*
 217: gb_est137:*
 218: gb_est138:*
 219: gb_est139:*
 220: gb_est140:*
 221: gb_est141:*
 222: gb_est142:*
 223: gb_est143:*
 224: gb_est144:*
 225: gb_est145:*
 226: gb_est146:*
 227: gb_est147:*
 228: gb_est148:*
 229: gb_est149:*
 230: gb_est150:*
 231: gb_est151:*
 232: gb_est152:*
 233: gb_est153:*
 234: gb_est154:*
 235: gb_est155:*
 236: gb_est156:*
 237: gb_est157:*
 238: gb_est158:*
 239: gb_est159:*
 240: gb_est160:*
 241: gb_est161:*
 242: gb_est162:*
 243: gb_est163:*
 244: gb_est164:*
 245: gb_est165:*
 246: gb_est166:*
 247: gb_est167:*
 248: gb_est168:*
 249: gb_est169:*
 250: gb_est170:*
 251: gb_est171:*
 252: gb_est172:*
 253: gb_est173:*
 254: gb_est174:*
 255: gb_est175:*
 256: gb_est176:*
 257: gb_est177:*
 258: gb_est178:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Description

C 1	97.8	1.5	569	102	A1813079	A1813079	20E11	Plat
C 2	85	1.3	905	245	A2550236	A2550236	ENT5E587T	Ent
C 3	83.4	1.3	1101	219	CNS00396	A1063921	Drosophill	Dros
C 4	81.8	1.3	849	245	A2546009	A2546009	ENT5E53TF	Ent
C 5	82.4	1.3	1101	219	CNS001T2	A10784714	ENT5E57T	Ent
C 6	80.6	1.3	641	236	A0946120	A0946120	Sheared E	Shear
C 7	80	1.2	877	245	A2531291	A2531291	ENTBQ34TR	Ent
C 8	79.4	1.2	749	238	A2129392	A2129392	OSNB010	OSNB
C 9	78.2	1.2	657	245	A0156164	A0156164	nbxdb00022E	nbxdb
C 10	78.2	1.2	900	245	A2549980	A2549980	ENTDP44TF	Ent
C 11	77.8	1.2	912	245	A2551092	A2551092	ENTFJ22TF	Ent
C 12	77.8	1.2	1101	219	CNS0021J	A2561936	Drosophill	Dros
C 13	77.6	1.2	1101	219	CNS012P8	A1101846	Drosophill	Dros
C 14	77.4	1.2	617	227	A0668185	A25486185	nbxdb00011	nbxdb
C 15	75.4	1.2	908	245	A2548467	A2548467	ENTEK30TF	Ent
C 16	74.4	1.2	860	219	CNS000U7	A10619453	Arabidops	Arabid
C 17	74	1.2	831	234	A0865895	nbdb00266T	ENT5E52T	Ent
C 18	74	1.2	974	219	CNS001TT	A1075423	Drosophill	Dros
C 19	73.6	1.1	1001	219	CNS0155H	A108536	Drosophill	Dros
C 20	72.2	1.1	1101	219	CNS011V2	A2551618	ENTDV54TR	Ent
C 21	71.2	1.1	843	245	A2551618	A2551618	ENTDV54TR	Ent
C 22	71	1.1	1101	219	CNS003B0	A1064078	Drosophill	Dros
C 23	70.6	1.1	1147	80	B273312	B1273312	GA_EB001	GA
C 24	70.4	1.1	605	227	A0367630	A0367630	nbxdb00021P	nbxdb
C 25	70.4	1.1	700	235	A0940248	A0940248	Sheared E	Shear
C 26	70	1.1	891	248	A2683582	A2683582	ENTKK47TF	Ent
C 27	70	1.1	963	219	CNS00M4L	A1054918	Drosophill	Dros
C 28	69.6	1.1	1025	219	CNS014J2	A11041216	Drosophill	Dros
C 29	69.4	1.1	739	174	B0127228	B0127228	EST473574	EST
C 30	69.2	1.1	813	219	CNS0119P	A1099988	Drosophill	Dros
C 31	69.2	1.1	910	219	CNS0113M	A10999775	Drosophill	Dros
C 32	69	1.1	867	219	CNS0054A	A1057618	Drosophill	Dros
C 33	68.6	1.1	561	230	A0577014	A0577014	nbxdb009000	nbxdb
C 34	68.6	1.1	619	230	A0576975	A0576975	nbxdb00900K	nbxdb
C 35	68.6	1.1	640	216	AMA96652	AMA96652	kb00408_Y	kb00
C 36	68.4	1.1	768	234	A0870251	A0870251	nbdb00366T	nbdb
C 37	67.8	1.1	536	258	TA281807Q	A1488498	T. brucei	T. brucei
C 38	67.8	1.1	1038	219	CNS0108N	A1098657	Drosophill	Dros
C 39	67.4	1.0	437	221	CNS030EVL	A1253269	Tetradodon	Tetradodon
C 40	67.2	1.0	1101	219	CNS000EVL	A10639206	Drosophill	Dros
C 41	67	1.0	951	247	A2676519	A2676519	ENTGV51TF	Ent
C 42	67	1.0	994	221	CNS04NOJ	A1298972	Tetradodon	Tetradodon
C 43	66.8	1.0	956	219	CNS0140V	AL104646	Drosophill	Dros
C 44	66.4	1.0	478	219	CNS000IT	AB091029	Arabidops	Arabidops
C 45	66.4	1.0	865	152	BG309498	BG309498	HVSMEC001	HVSMEC001

ALIGNMENTS

Description

A1810379 20E11 Pin
A2550256 ENTFWS8TR
AL0663921 Drosophila
A2546009 ENTFWF3TF
AL078714 Drosophila
A0946120 Sheared D
A2531291 ENTBQ24TR
A123292 OSNB5010
AQ156164 nbxb0002H
A254980 ENTD94TF
A2551092 ENTFJ22TF
A0661936 Drosophila
AL101846 Drosophila
A0368185 loxbD001L
A2348467 ENTERK07R
A0191453 Arabidops
A0865895 nbex0026F
AL075432 Drosophila
AL105023 Drosophila
AL108536 Drosophila
A2551618 ENTDF54TR
AL064078 Drosophila
B1273312 GA_Ebd01
A0367630 loxbD002D
A0940248 Sheared D
A2683582 ENTKK47TR
AL054918 Drosophila
AL104216 Drosophila
BG127728 Bst473374
AL099988 Drosophila
AL099775 Drosophila
AL057618 Drosophila
A0577014 nbxb0090O
A0576975 nbxb0090K
AM466652 kp04008.Y
X0870251 nbe0036P
AL488498 T. brucei
AL084657 Drosophila
AL053269 Tetradodon
AL069706 Drosophila
A2676519 ENTGWS1TF
AL298972 Tetradodon
AL104646 Drosophila
AD091029 Arabidops
AG309498 HVMSC001

Forest Biotechnology Group
North Carolina State University
Dept. of Forestry, NC State University, 6113 Jordan Hall, Raleigh
NC, 27695-8008
Tel: 919-515-7800
Fax: 919-515-7801
Email: rosswhet@unity.ncsu.edu
Seq primer: T3.

FEATURES
source Location/Qualifiers
1..569
 /organism="Pinus taeda"
 /db_xref="taxon:3352"
 /collection="Pine Lambda Zap Xylem library"
 /rna_type="differentiating xylem"
 /note="Vector: Lambda Zap; Site: 1; EcoRI; Site 2: XhoI;
Differentiating xylem was collected from the main stem of
a 35-year old loblolly pine tree harvested during the
growing season. RNA isolation and library preparation
followed the methods of Allona et al., PNAS 95:9693-8,
1998"

BASE COUNT 206 a 74 c 120 g 169 t
ORIGIN

Query Match 1.5%; Score 97.8; DB 102; Length 569;
Best Local Similarity 48.5%; Pred. No. 1.3e-10;
Matches 270; Conservative 0; Mismatches 287; Indels 0; Gaps 0;

QY 4730 actaatgaattatgcacaataatagtttgtaattgaatgaacaaatcaacttagac 4789
|| || || || || || || || || || || || || || || || || || || || || || ||
Db 13 ACAATTGATACGATTCGCTTGGCTTAGCCACAAMAACNTATTGGAAACTACATCATGTGAT 72

QY 4790 gtcgagtcggcgtatctcaaatgcctctatactcaatccaatccaatatgatgttctct 4849
|| || || || || || || || || || || || || || || || || || || || || || ||
Db 73 GTCAATCTGCTTTCTTAATAGTGATTTAAGGAAGAAGTATTATTGTGTTCAACCTGAA 132

QY 4850 ccctaactgaactcttgaaagaanaaacatcttgtttatitigaacgctctgtcatagg 4909
|| || || || || || || || || || || || || || || || || || || || || || ||

Db 133 GGCGTTTGCAACACGGGACACAGGACATCTGTTTGCAAGGTAAAGAAAGCATTGTATGCT 192

QY 4910 ttaaacaagtcggglttggaaagtatcacacactcaaaaagatattggaagatcatttgt 4969
|| || || || || || || || || || || || || || || || || || || || || || ||
Db 193 TTGGAACACGGCACCAAGCATGTATGTATTAAGCAATTTGATACCTCTTCTTCAGATGGA 252

QY 4970 ttlaactaagtttaacaaatgtgttatctcaactgaatgaatataagaagatcagta 5029
|| || || || || || || || || || || || || || || || || || || || || || ||

Db 253 TTTGTGAAGAAGCAAAAATGATCTCACTTATTATGTTTAAAGAAATATAAAGAAATGTT 312

QY 5030 atatatagtttatatgtttgaatgatbatcctatgtttggaagttcacaaaaagttatt 5089

Db 313 GCCTTATATCTTGTATGTGATGATCTCATTAATAACAGGATGCTGCACAAAGTTAAT 372

RESULT	1
LOCUS	A1813079
DEFINITION	A1813079 569 bp mRNA EST 08-Jul-1999 20E11 Pine Lambda Zap Xylem library Pinus taeda cDNA, mRNA
ACCESSION	A1813079
VERSION	A1813079.1 GI:5424175
KEYWORDS	EST.
SOURCE	loblolly pine. <i>Pinus taeda</i>
ORGANISM	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Coniferales; Pinaceae; Pinus; Pinus
REFERENCE	1 (bases 1 to 569)
AUTHORS	Whetten,R.W., Kinlaw,C.S., Retzel,E. and Sederoff,R.R.
TITLE	The Pine Gene Discovery Project
JOURNAL	Unpublished (1999)
COMMENT	Contact: Ross Whetten

FEATURES		Location/Qualifiers			
source		1..569			
		/organism="Pinus taeda"			
		/db_xref="taxon:3352"			
		/clone_lib="Pine Lambda Zap Xylem library"			
		/issue_type="differentiating xylem"			
		/note="Vector: lambda Zap; Site 1: EcoRI; Site 2: XhoI; a 35-year old loblolly pine tree harvested during the growing season. RNA isolation and library preparation followed the methods of Allona et al., PNAS 95:9693-8, 1998"			
BASE COUNT		206 a	74 c	120 g	169 t
ORIGIN					
Query Match 1.5%, Score 97.8; DB 102; Length 569;					
Best Local Similarity 48.5%; Pred. No. 1.3e-10;					
Matches 270; Conservative 0; Mismatches 287; Indels 0; Gaps					
QY	4730	actataagatattatcacataatagtggtgtaattagaatgacaattcaacatttagac	4769		
DB	13	ACATTTAGATTTCATTCATTGTCCTTACCCCAAAACATTAATTGGAAACATCATCATTTGGAT	72		
QY	4790	gtcgagtcgagcgatcctaatagtccctctaactcaatcaaalccaattatgcttcct	4849		
DB	73	GTCAATCTGCTTTCTPAAATGGTGATTTAAGAGAAAGTTATTGGTTCAACCTGAA	132		
QY	4850	cctaatacgtacaccttgaagaagaaacacatggttggtatgaaagttcgtctatggg	4909		
DB	133	GGGTTTGTCAAAACAGGACAGAGAGCATGTGGTTTGCAGGTTAAGAAAGCATGTGATGCT	192		
QY	4910	ttaaacagtcgaggttbgaaatggtcacactacaactcaagaagatitggaagatggt	4969		
DB	193	TTGAACACGAGCACCAGGTCATGATGTAAGATTTGATGTAACCTCTTCTTCACAGATGA	252		
QY	4970	ttlactcaagtttacaacatgagtgttattlccaactlgaatalygaagaggaacgta	5039		
DB	253	TTTGTGAAGAACCAAAATGATCTCFACCTTGATGTTAAGAAAGTAAAGAAATGTT	312		
QY	5030	ataatattaggtttatagtgatgatatcttattggttgaagtccaacaaagtatt	5089		
DB	313	GCTTTAATATCTTGTATGTGATGATGATCTCATTTAACAAGGTAGTCTCAACAAGTTAATT	372		
QY	5090	gataattlgtgacaaatgagagatcaatttgaagtgaagtggttvgtgaataatca	5149		
DB	373	GAAAGAAATTAAGTTCACTGTGACACAGGAATTTGAAATGAAGAAGTTGGTGGAATGAT	432		
QY	5150	aattactctgtatitgaattcgtlaaaacgcgaactcgttatattatctcaagaaaa	5209		
DB	433	TACTGTTTAAAGAAATTTGAGTATGGAGAACACCTGGTAAACCTTTGATCAGTCAAGAGCA	492		
QY	5210	tttcacaagaatattactaaagattccaactagatgacatataatgggaanaacatccc	5269		
DB	493	TATACAAAGGAGATTTCCTAAAGAAATGTCATATGACAGATTGCCAAAGCTATGTCAACTCCC	552		
QY	5270	tgagattccgaatgaca	5286		
DB	553	TTAGAACGAGATGCAAA	569		

genomic, DNA sequence.
 accession 42550256
 version 42550256.1 GI:1175557
 keywords GSS.
 source Entamoeba histolytica.
 organism Entamoeba histolytica
 Eukaryota; Entamoebidae; Entamoeba.
 reference 1 (bases 1 to 905)
 authors Loftus,B., Van Aken,S. and Fraser,C.
 title Determination of clone end sequences from Entamoeba histolytica
 HMI:IMSS sheared DNA library
 journal Unpublished (2000)
 comment Contact: Brendan J Loftus
 department of Eukaryotic Genomics
 the Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 tel: 301 838 0200
 fax: 301 838 0208
 email: b1loftus@igr.org
 clones are derived from the Entamoeba histolytica HMI:IMSS sheared
 DNA library
 seq primer: M13-Reverse
 class: shotgun
 high quality sequence start: 100
 high quality sequence stop: 872.
 location/Qualifiers
 1..905

/clone.lib="Entamoeba histolytica Sheared DNA"
 /note="Vector: pHOSt; Site_1: Bst I; Constructed at The
 Institute for Genomic Research (TIGR), Rockville, MD.
 Genomic DNA isolated from broth cultures of E. histolytica
 using a method described by Clark and Diamond (Clark,
 C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
 method for isolate identification. Exp. Parasitol.
 77:450.). The DNA was mechanically sheared to give a
 tight size distribution (~2 kb). The v + i method used for
 the library construction is described in detail in Smith,
 H.O. and Venter, J.C. (Making small insert libraries for
 whole genome shotgun sequencing projects. In Genome
 Sequencing: A Practical Approach, eds. M. Vaubin and B.
 Barrell, Oxford University Press, 1999)."
 203 a 260 c 46 g 396 t

Query Match	1.3%	Score	85	DB	245	Length	905
Best Local Similarity	46.7%	Pred	No.	8.1e-08			
Matches	305	Conservative	0	Mismatches	345	Indels	3
						Caps	1

QY	3562	agatcattcggcgtgcgaactgtttaagaacaagctcgaagaattgcatcttcga	3621
Db	715	AGACGAGAGATGAAGAAGAGCTGTGATGTGACGAAGAAGAAAGAAATGATGCA	656
QY	3622	ctaccagaacatcagtcgcgaagatgcaacagcaaatccctcgttactcgtgctaataagaa	3681
Db	655	TGAATTTCAATTTCAGAGATGCAAGAGTGAAGATGATGTATGATGAAGAAGATGATGATGA	596
QY	3682	tcgggtactcgaagaatagatagaggagaanaatttcatttcgcgggggtgcgtcga	3741
Db	595	TGAGATGATGATATGAAGATGAGAGAGCTGTGATGTATGAAAGAAAGCAGAAAGAAAGATGA	536
QY	3742	tctcgtcgtgactcaactcaaatglttgagcaactaaatglttgaaacagagatcgtgtaa	3801
Db	535	TGATGATCAATTTCAATTAGACAGTGAAGAGATGAAGAAGATGAAGAAGATGATGATGA	476
QY	3802	cagtcacaattcgaagcagaacttttcgaaggagagaataactaaatgaacaacatcgatat	3861
Db	475	TGATGATCAAAATATATGAGCTGTGTAAGATGATGAAAG---TGATGCAAGATGATGAAAGA	419
QY	3862	agttgatactcgttgctaaagttaatttgagaatgaaaaaatctctccatlaattcaattaga	3921

[illegible]

RESULT	3
CNS00396/c	
LOCUS	
DEFINITION	CNS00396 1101 bp DNA GSS 03-JUN-1999
ACCESSION	Drosophila melanogaster genome survey sequence TERN3 end of BAC #
VERSION	BACR08k10 of RPTC-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
KEYWORDS	AL063921 AL063921.1 GI:4941778
SOURCE	GSS:
ORGANISM	fruit fly.
REFERENCE	Drosophila melanogaster.
TITLE	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
JOURNAL	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
	Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
	1 (bases 1 to 1101)
	Genoscope.
	Direct Submission
	Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage ;

JOURNAL
Submitted (02-JUN-1999) Genoscope - Centre National de Séquençage
BP 191 91006 EVRY cedex - FRANCE (E-mail : secref@genoscope.cns.fr)
- Web : www.genoscope.cns.fr

COMMENT

Determination of this BAC end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org/TheBDGP/Drosophila_melanogaster_BAC_library was prepared by Kaitoyo Oosagawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named PCR1-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain Y2: cn bw sp, the same strain used for the BDGP's pl and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/Drosophila_bac.htm.

FEATURES	source	Location/Qualifiers
		1. 1101
		/organism="Drosophila melanogaster"
		/db_xref="taxon:7227"
		/clone_id="RPCT-98"
		/clone="BACR08K10"
		/note="end : TE13"
BASE COUNT	201 a	64 c 131 g
ORIGIN		202 t 503 others

Query Match	13%	Score 83.4	DB 219	Length 1101
Best Local Similarity	17.3%	Pred. No. 1.9e-07		
Matches 118	Conservative 289	Mismatches 274	Indels 0	Gaps 0

```

QY 28 gaactgatacaatgaatgatgtattatgaagaattgcaaaattttccatcacacac 87
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1100 KARRMGDDTMDRDRKDDMTKMWKMRADRRMAGDADWAMDGAGTWTWTW 1041
QY 88 aggtgatgacagaactaaactatgtgtgtatataaataaggtatgaataccaacat 147
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1040 MWWWATWTDWMDKMMWMTATATWMTWAMPADNAGRGAGKRRDRAATDADA 981
QY 148 cccagaatatacaacagatagaaggaaggtttcaatatatactcttgaataaact 207
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 980 GRDDGKRRKDKDKDKDDGDKGKKAKAKAKAKAKAKAKAKAKAKAKAKAK 921
QY 208 tgcgttaactacatacaacagatgacagtcacacgcacacacaggttaagaag 267
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 920 DDDGAGDKDDGDKGADDDDTGTDDDDDKMDWDMDAKGWTGATWAMAATDMMW 861
QY 268 ttatattcaatcagatagaagtcagatgataatcaatctgcctcaaatagcgtgt 327
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 860 MADADWMTDAADWADWMDWMDWMDWMDWMDWMDWMDWMDWMDWMDWMDWMD 801
QY 328 ataattcagctcagatgtgtatgtatgtatgtatgtatgtatgtatgtatgtat 387
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 800 RRADDKRRADRRDRAATWTTTDTTDDDKWKTDTWTRWADRTWDDDDDDDR 741
QY 388 attcaacagatgagtcgcaagaatgatgatacagaagaaggaagtcagatgaagt 447
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 740 AGTAGRRKWRTRKRRKRRDTRMDADADDTARDDRRRRGDDGADGKKTGRKRR 681
QY 448 tgaatcagctatgctatagtaagtgatgaatgaacatatacagagtcagatgaat 507
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 680 RATWTDWMDWMDWMDWMDWMDWMDWMDWMDWMDWMDWMDWMDWMDWMDWMD 621
QY 508 ttcatataaatttaagtaactcagatgagtcgagtcgagtcgagtcgagtcgagtc 567
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 620 KMDWKTARRDRWMDWMDWMDWMDWMDWMDWMDWMDWMDWMDWMDWMDWMDWMD 561
QY 568 atggaatgaagaatttaatttcaatcagtcgagtcgagtcgagtcgagtcgagtc 627
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 560 TWTWTAARAAMWMAWMTATTTTWTWTTTWTWTTTWTWTTTWTWTTTWTW 501
QY 628 tgactataatccaaagaatgaagtcagtcagtcagtcagtcagtcagtcagtcag 687
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 500 AAWTAAAMWMAWMAWMTATTTTWTWTTTWTWTTTWTWTTTWTWTTTWTW 441
QY 688 ttggtgttacaatgtgtt 708
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 440 TTTWTTWMAATTTATTTWT 420

```

```

RESULT 4
LOCUS A2546009 849 bp DNA GSS 14-NOV-2000
DEFINITION ENFM53TF Entamoeba histolytica Sheared DNA Entamoeba histolytica
ACCESSION A2546009
VERSION A2546009.1 GI:11167130
KEYWORDS GSS.
SOURCE Entamoeba histolytica.
ORGANISM Entamoeba histolytica.
REFERENCE 1 (bases 1 to 849)
AUTHORS Loftus,B., Van Aken,S. and Fraser,C.
TITLE Determination of clone end sequences from Entamoeba histolytica
JOURNAL HMI:IMSS sheared DNA library
COMMENT Unpublished (2000)
Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: b.loftus@tigr.org

```

Clones are derived from the Entamoeba histolytica HMI:IMSS sheared DNA library
 Seq primer: M13-Forward
 Class: shotgun
 High quality sequence start: 26
 High quality sequence stop: 796.

FEATURES

source
 1..849
 /organism="Entamoeba histolytica"
 /strain="HMI:IMSS"
 /db_xref="taxon:5759"
 /clone_lib="Entamoeba histolytica Sheared DNA"
 /note="Vector: pHD31; Site:1; Bst I; Constructed at The Institute for Genomic Research (TIGR), Rockville, MD. Genomic DNA isolated from broth cultures of E. histolytica using a method described by Clark and Diamond (Clark, C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a method for isolate identification. Exp. Parasitol. 77:450.). The DNA was mechanically sheared to give a tight size distribution (~2 kb). The v + i method used for the library construction is described in detail in Smith, H.O. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaubin and B. Barrell, Oxford University Press, 1999)."

BASE COUNT 199 a 236 c 43 g 371 t
 ORIGIN

Query Match 1.3%; Score 82.4; DB 245; Length 849;
 Best Local Similarity 46.7%; Pred. No. 2.9e-07;
 Matches 297; Conservative 0; Mismatches 336; Indels 3; Gaps 1;

```

QY 3539 gagagatcaaatgtctggaacagatcattcgtgtcgcagtcgttgaagaacagtc 3598
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 697 GAAGCAATGATATGAGAGAGAGATGATGATGATGATGATGATGATGATGAT 638
QY 3599 tcagaagatgctactctcagctcagctcagctcagctcagctcagctcagctcag 3658
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 637 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 578
QY 3659 cccctgtactgctaaatgaagatcggtaactggaataaataagatgaggaataat 3718
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 577 GAAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 518
QY 3719 tcatttcggggggtgatgatgatgatgatgatgatgatgatgatgatgatgat 3778
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 517 GAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 458
QY 3779 gtgaaacagagagatgctggtgaacagtcgaatcgaagaagttctcaagaaggaga 3838
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 457 GAAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 398
QY 3839 atacttaatgacaacatgatatagttgactgctgctaaagttaattgagaatga 3898
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 397 GAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 338
QY 3899 atcttccttaattcaatagatgatactcagactgtgtcagactcagactcagact 3958
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 337 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 278
QY 3959 agcaatcaacagatcgcagatcgaatcgaatgaatgaatgaatgaatgaatgaat 4018
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 277 ---GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 221
QY 4019 aaaaatcgtgaataatccaaacacatgaaagtatccctgtcgtatagaagatgat 4078
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 220 GAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 161
QY 4079 gaattgaaacgataatgtgtgaatgagaatgagatgaatgaatgaatgaatgtgt 4138
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 160 GAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 101

```

Oy	4139	gaagcttcacccacgataatgatgtc 4174
Dd	100	GAAATGATCATGAAGATGATCAACATCCAGAATT 65
RESULT 5	CNSOULT2	
LOCUS	CNSOULT2	1101 bp DNA GSS 14-JUN-1999
DEFINITION	Drosophila melanogaster genome survey sequence TET3 end of BAC: BACR4BP19 of Rpci-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.	
ACCESSION	AL078714	
VERSION	AL078714.1	GI:5102004
KEYWORDS	GSS.	
SOURCE	fruit fly.	
ORGANISM	Drosophila melanogaster Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila. 1 (bases 1 to 1101)	
REFERENCE	Genoscope. Direct Submission Submitted (11-JUN-1999) Genoscope - Centre National de Sequencage BP 191 91006 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr Web : www.genoscope.cns.fr) Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACS. For further information please see http://www.fruitfly.org/The_BDGP_Drosophila_melanogaster_BAC_library.html prepared by Kazutoyo Osogawa and Aaron Mammeter in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named Rpci-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain Y2: cn bw sp, the same strain used for the library pl and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.	
FEATURES	source locationQualifiers	1..1101 /organism="Drosophila melanogaster" /db_xref="taxon:7227" /clone_1ib="RPCI-98" /clone="BACR4BP19" /note="end : TET3"
BASE COUNT	469 a 6 c 69 g 151 t 406 others	
ORIGIN		
Query Match	1.3%; Score 81.8; DB 219; Length 1101;	
Best Local Similarity	29.2%; Pred. No. 4.1e-07;	
Matches 212; Conservative 166; Mismatch 344; Indels 3; Gaps 1		
Oy	1177	aagcaccaatatgaacttaatttgtagttaccaggccagagaacaacaagg 1236
Dd	273	AAA 332
Oy	1237	aaaacagaggaagactctcggaacaattcaaagaagacctgaatcaagaagsg 1296
Dd	333	AAA 392
Oy	1297	aatatgaagaatccaatcgatagataaagtgtaaagaagaagaagaagaaaa 1356
Dd	393	AAA 452
Oy	1357	aacttcaactggaaaacaagagtgcgtcttatctaattgtyttaatacatataa 1416
Dd	453	AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATTTTTWTWTKKATTTTWTDDA 512
Oy	1417	ttagcgcaaaacacagtttcagtagaanaattcattctcttaatgctcttgtgaagt 1476

D	b	513	ADKNAAMWDITTTTANMTWTWNTATATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTAKKARDTWTKKAATAKT	572
O	y	1477	aatgaatttaaaaggtttaaggtttaa---caagtactcagtgatgatactgcyggccac	1533
D	b	573	AAAAAAMTTTTTTTAATTTTWTCTTTAKRTWAMKMAMDATTTTWTWTWTWTWTWT	632
O	y	1534	aaractcgttcgagaacaataagaatatatgcgatcgtaagcagaacgaatgaagt	1593
D	b	633	AATTTTWTWTWTWTAAVTTTTTTTTTTTTTTTAKTTTTTTTTTTTTATTAATAAATAAAAWTTDTIW	692
O	y	1594	tctcgttcgtcgtatgctcacattagaacagatgtgatgigtatctaattaacagat	1653
D	b	693	AAAMWTTTTKKKKKKAAADAKMKDKAMGAKKATKTCKKKKAAAMAADKKDRKKGGKKK	752
O	y	1654	cggatcttctcgtacgtacgttagaataacatctgtatttaccagaagtccttatct	1713
D	b	753	GKKKKKKKGKKKKKKKKGGWKGGKKAGDDADKDTTKKKKKRATTTTKKKGGKK	812
O	y	1714	tatgagcttgaacaacaattggaacagcggatllaatgcttactactaagaacagt	1773
D	b	813	KKRAAKKKKADTTKTTKWDAAAAAAAAAAKRTKDGGKKKKTKTTKXXXXXXXXXXGGK	872
O	y	1774	gattgatattaaaccaaatgtygtcctcactattatiglytcitcaaagaatgctcgt	1833
D	b	873	KDAAKKKKKCGTCKKKKKKKGGGKGGKGGKKDDDAANAARKKKKGGGSGKKK	932
O	y	1834	tcttattatggtcctcaattcagtatgaagaatctttagaatgattgattatagatg	1893
D	b	933	KGGKKGGKGGKKRKTKTKTKKRRKKKKADAAKGTKKRRAAAADAADWTDAATKRRKKK	992
O	y	1894	tttgg 1898	
D	b	993	KDKKK 997	

RESULT	6	
LOCUS	AQ0946120/c	GSS 27-JAN-2000
DEFINITION	Sheared DNA-46J23.TR Sheared DNA Trypanosoma brucei genomic clone	
ACCESSION	AQ0946120	
VERSION	AQ0946120.1	GI:6769385
KEYWORDS	GSS.	
SOURCE	Trypanosoma brucei.	
ORGANISM	Trypanosoma brucei. Eukaryota; Euzlenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.	
REFERENCE	1 (bases 1 to 641) El-Sayed,N., Zhao,S., Zhao,H., Gill,S., Suh,E., Malek,J., Fujii,C., Gerard,C., Leech,V., de Jong,P., Ullu,E., Melville,S., Donelson,J., Fraser,C. and Adams,M. Determination of clone end sequences from Trypanosoma brucei GUTat 10.1 sheared DNA library Unpublished (1999)	
TITLE	Other_GSS: Sheared DNA-46J23.TF	
JOURNAL	Contact: Najib M. El-Sayed Department of Eukaryotic Genomes The Institute for Genomic Research 9712 Medical Center Dr., Rockville, MD 20850, USA ; Tel: 301 838 0200 Fax: 301 838 0208	
COMMENT	Email: nelsayed@tigr.org Clones are derived from the "Trypanosoma brucei GUTat 10.1 sheared DNA library constructed at TIGR. Clones will be available for distribution through ATCC. Sheared DNA end sequences search page: http://www.tigr.org/tdb/mbd/tbdb/ . Seq primer: MJ3-Reverse Class: Shocgun.	
FEATURES	Location/Qualifiers	
source	1..641	
	/organism="Trypanosoma brucei" /strain="TREU927/4 GUTat 10.1" /db_xref="taxon:5691"	

QY 3959 agcaatcaacgaatccacatcattcattcgaataatataccagtgattag 4018
 DB 386 GATGAAGATGATGAAGAGATGATGAAGAGATGATGAAGATGATGAAGAG 327
 QY 4019 aaaaactgaataatccaaaacacattgaagatcctctgctgataagagattgag 4078
 DB 326 GAAGAAGATGATGATGAAGAGATGATGATGAAGATGATGATGAAGAGAT 267
 QY 4079 gaattgaacgtataatgttgataatgagattgagattgagattgagattgag 4138
 DB 266 GAAGATGATGATGAAGAGATGATGATGAAGATGATGATGAAGATGATGAAG 207
 QY 4139 gaagctaacctaccagatgaataatgattgattgattgattgattgattgatt 4198
 DB 206 GAAGAAGATGATGATGAAGAGATGATGATGAAGAGATGATGATGAAGATGATGA 147
 QY 4199 aatcatagatgtcagagagaaagaaatacacattgagaaagattacagaaaaa 4258
 DB 146 GACGAGATGATGATGAAGAGATGATGAAGAGATGATGAAGAGATGATGAAGAG 87

RESULT 8
 A2129392/c

LOCUS A2129392 749 bp DNA GSS 02-JUN-2000
 DEFINITION OSJNB0101G09f CUGI Rice BAC Library (EcoRI) Oryza sativa genomic
 clone OSJNB0101G09f, DNA sequence.

ACCESSION A2129392
 VERSION A2129392.1 GI:8207302
 KEYWORDS GSS.
 SOURCE Oryza sativa.
 ORGANISM Oryza sativa

REFERENCE 1 (bases 1 to 749)
 WING, R.A. and DEAN, R.A.
 A BAC End Sequencing Framework to Sequence the Rice Genome
 Unpublished (1998)
 CONTACT: Wing RA
 CLEMSON UNIVERSITY GENOMICS INSTITUTE
 100 JORDAN HALL, CLEMSON, SC 29634, USA
 TEL: 864 656 7288
 FAX: 864 656 4293
 EMAIL: rwing@clermson.edu
 SEQ PRIMER: G7AAACGACGCGCCAGTG
 CLASS: BAC ends
 HIGH QUALITY SEQUENCE START: 58
 HIGH QUALITY SEQUENCE STOP: 578.
 LOCATION/Qualifiers

FEATURES
 source

/organism="Oryza sativa"
 /strain="Japonica"
 /cultivar="Nipponbare"
 /db_xref="taxon:4530"
 /clone="OSJNB0101G09f"
 /clone_lib="CUGI Rice BAC Library (EcoRI)"
 /issue_type="leaf"
 /lab_host="E. coli DH10B"
 /note="Vector: pBACindigo; Site_1: EcoRI; Site_2: EcoRI;
 Rice is the most important food crop in the world. Half of
 the world population, especially those inhabiting highly
 populated areas of the humid tropics and subtropics, rely
 on rice as their primary source of carbohydrate.
 Monocotyledonous rice is a diploid plant (2n=24) with a
 haploid genome equivalent of 431 Mbp (Arumuganathan and
 Earle, 1991). The relatively small genome of rice, three
 times larger than that of Arabidopsis, makes it suitable
 for genomic studies. In order to facilitate positional
 cloning, physical mapping and genome sequencing of rice,
 we have constructed a BAC library from Oryza sativa.
 Nipponbare variety using EcoRI as the cloning enzyme. The
 library contains 55,296 clones with an average insert size

of 121 Kb providing approximately 15 haploid genome
 equivalents. The deep coverage allows the isolation a
 particular sequence with a probability of 99.9 %. Three
 high density filters, each containing 18,432 clones
 (doubly spotted), represent the whole library for colony
 screening and can be requested from the Clemson University
 BAC/EST Resource Center (www.genome.clemson.edu)."
 BASE COUNT 252 a 162 c 126 g 203 t 6 others

Query Match 1.2%; Score 79.4; DB 238; Length 749;
 Best Local Similarity 52.8%; Pred. No. 1.2e-06;
 Matches 195; Conservative 0; Mismatches 171; Indels 3; Gaps 1;

QY 4861 acccttgaagaacacacattgttgattgaacgctctgctatggttgaacagtc 4920
 DB 633 ATCTCTAACCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 574
 QY 4921 gggattgaaatgatacacaacacacacacacacacacacacacacacacacacac 4980
 DB 573 AACCAAGGCGCATGATATTCACAGCTGAGTGAAGAACTGTGGAATGCTT---TGAGC 517
 QY 4981 ttacacaaatgattgttatttcaacattgaatgaagagagatcagtaataattag 5040
 DB 516 TTCAAAGCGCGATCTTCAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 457
 QY 5041 ttataatgattgatttatttcttattgattgattgattgattgattgattgattgatt 5100
 DB 456 GGTATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 397
 QY 5101 ggaatcaatggaatcatttgaagattgaagattgattgattgattgattgattgatt 5160
 DB 396 GAAGATATTAACAAGAGGATTTGCACTTAAGATTGGTACCTGATTAATTAATTAAT 337
 QY 5161 tattgaatttgcataaaccggaatcgttatttatttatttatttatttatttatttatt 5220
 DB 336 AATTGAGGTAACTAAAGTTTCAATGATTAATTAATTAATTAATTAATTAATTAATGA 277

QY 5221 attactaa 5229
 DB 276 TTTATTAA 268

RESULT 9
 A0156164 657 bp DNA GSS 12-SEP-1998

LOCUS nbxb0002H20r CUGI Rice BAC Library Oryza sativa genomic clone
 DEFINITION nbxb0002H20r, DNA sequence.

ACCESSION A0156164
 VERSION A0156164.1 GI:3552253
 KEYWORDS GSS.
 SOURCE Oryza sativa.
 ORGANISM Oryza sativa

REFERENCE 1 (bases 1 to 657)
 WING, R.A. and DEAN, R.A.
 A BAC End Sequencing Framework to Sequence the Rice Genome
 Unpublished (1998)
 CONTACT: Wing RA
 CLEMSON UNIVERSITY GENOMICS INSTITUTE
 100 JORDAN HALL, CLEMSON, SC 29634, USA
 TEL: 864 656 7288
 FAX: 864 656 4293
 EMAIL: rwing@clermson.edu
 SEQ PRIMER: GGAACACGCTATGACCATG
 CLASS: BAC ends
 HIGH QUALITY SEQUENCE STOP: 235.
 LOCATION/Qualifiers

FEATURES
 source

/organism="Oryza sativa"
 /strain="japonica"
 /cultivar="Nipponbare"
 /db_xref="taxon:4530"
 /clone_lib="Oxg Rice BAC Library"
 /issue_type="Leaf"
 /lab_host="E. coli DH10B"
 /note="Vector: pRebAC11; site:1: HindIII; site:2: HindIII; Rice is one of two most popular grains in the world. Half of the world population especially those inhabiting highly populated areas of the humid tropics and subtropics, rely on rice as their primary source of carbohydrate. Monocotyledonous rice is a diploid plant (2n=24) with a haploid genome equivalent of 431 Mbp (Arumuganathan and Earle, 1991). The relatively small genome of rice, three times larger than that of Arabidopsis, makes it suitable for genomic studies. In order to facilitate positional cloning, physical mapping and genome sequencing of rice, we have constructed a BAC library from Oryza sativa, Nipponbare variety. The library contains 36,864 clones with an average insert size of 128.5 Kb providing 10.9 haploid genome equivalents. The deep coverage allows the isolation a particular sequence with a probability of 99.9%. Two high density filters, each containing 18,432 clones (doubly spotted), represent the whole library for colony screening."

BASE COUNT 197 a 99 c 139 g 221 t 1 others
 ORIGIN

Query Match 1.2%; Score 79; DB 224; Length 657;
 Best Local Similarity 50.0%; Pred. No. 1.5e-06;
 Matches 224; Conservative 0; Mismatches 221; Indels 3; Gaps 1;

OY 4899 ctctcattggttaaaacagctgggttggaatggtatcacactacaaaggtttg 4958
 |||||
 DB 7 CTTTATATGCTTAAACAGCGACCTAGAGCTTGTACTCTAGCTTACAGCAAGTTAT 66
 OY 4959 aagacattggtttactcaagtttacaatgatgttatttcaatgaaatcgaag 5018
 |||||
 DB 67 ATGATATGGGAT---CCAAAGGCTCCAGAGCTGATCTCTCTTTTAAACAAAG 123
 OY 5019 agggatcagaaatattaggtttatagtgtatgatatcttctatggttgaagtcac 5078
 |||||
 DB 124 GAGATTGATATATTTGATCTCATTTATGATGATATATTTGACACAGTTCTAGAC 183
 OY 5079 aaaaattatgataattttgtgacaaatgagagatcatatttgaagttaagtgtg 5138
 |||||
 DB 184 AGGAAGCTGTTCACGCCCTTCCTCGGAGTTTACAAAGAGTTTCTCTGAAAGATCTTG 243
 OY 5139 gtgaatatacaattatcttgatgaaatctgtaaaacgaatctgttattat 5198
 |||||
 DB 244 GAGATCTTCTACTTCTCTGATGATGATGATGATGATGATGATGATGATGATGAT 303
 OY 5199 ctcaaaaataatttctcaagaatctcaagaatctcaagaatctcaagaatctcaaga 5258
 |||||
 DB 304 CTCAAGATTAATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 363
 OY 5259 aaaaatcaccctgattccgaatgacaataatgaagaagttcaataatctgttgaaga 5318
 |||||
 DB 364 TTAGTACACCCCTGCTTACAGAGCAAAATTTGCTTAAATAGAGACATCTTTAGAC 423
 OY 5319 ttaatcagaagaatttttgaagaagt 5346
 |||||
 DB 424 CAAATGATGACACACATTTAGAGTGT 451

RESULT 10
 AZ549980 900 bp DNA GSS 14-NOV-2000
 LOCUS ENTDD94TF Entamoeba histolytica Sheared DNA Entamoeba histolytica
 DEFINITION genomic, DNA sequence.

ACCESSION AZ549980
 VERSION AZ549980.1 GI:1175122
 KEYWORDS GSS.
 SOURCE Entamoeba histolytica.
 ORGANISM Entamoeba histolytica
 Eukaryota: Entamoebidae: Entamoeba.
 REFERENCE 1 (bases 1 to 900)
 Loftus, B., Van Aken, S. and Fraser, C.
 Determination of clone end sequences from Entamoeba histolytica
 HMI:IMSS sheared DNA library
 Unpublished (2000)
 JOURNAL COMMENT
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: b.loftus@ligr.org
 Clones are derived from the Entamoeba histolytica HMI:IMSS sheared
 DNA library
 Seq primer: M13-Forward
 Class: Shotgun
 High quality sequence start: 20
 High quality sequence stop: 890.
 Location/Qualifiers
 1..900
 /organism="Entamoeba histolytica"
 /strain="HMI:IMSS"
 /db_xref="taxon:5759"
 /clone_lib="Entamoeba histolytica Sheared DNA"
 /note="Vector: pHO51; Site:1: Bst I; Constructed at The
 Institute for Genomic Research (TIGR), Rockville, MD.
 Genomic DNA isolated from broth cultures of E. histolytica
 using a method described by Clark and Diamond (Clark,
 C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
 method for isolate identification. Exp. Parasitol.
 77:450-). The DNA was mechanically sheared to give a
 tight size distribution (~2 kb). The v + i method used for
 the library construction is described in detail in Smith,
 H.O. and Venter, J.C. (Making small insert libraries for
 whole genome Shotgun sequencing projects. In Genome
 Sequencing: A Practical Approach, eds. M. Vaudin and B.
 Barrell, Oxford University Press, 1999)."

BASE COUNT 376 a 63 c 235 g 226 t
 ORIGIN

Query Match 1.2%; Score 78.2; DB 245; Length 900;
 Best Local Similarity 45.4%; Pred. No. 2.3e-06;
 Matches 281; Conservative 0; Mismatches 338; Indels 0; Gaps 0;

OY 3582 ttgttaagaacacgtctcagaagatggtgcatctctgactaccocgaactagtcgt 3641
 |||||
 DB 227 TTGAATTTGAAAGACACTTTTAAATGATATGACTATGAAATTTGAAACATGAAAGATT 286
 OY 3642 aagatgaacagcaaatccctcgtactgctaaatgaagatcgggtgaactgaataaataag 3701
 |||||
 DB 287 TAGATGACGAGAAATTTGATAGATGAAAGATGAAAGATGAAAGATGAAAGATGAAAG 346
 OY 3702 atgaagggaataatttcaattcccgggggtgatgatgattcgtcgtgataactcaa 3761
 |||||
 DB 347 AAGATGAAGACGAGACGATGATGAGACGATGAGATGAGATGAGATGAGATGAGATGATG 406
 OY 3762 atgttgacaatcattgttgaacagagatgctgttaacagccatcaagagagaag 3821
 |||||
 DB 407 AAGATGATGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 466
 OY 3822 ttctcaaggaggaataacttaattgaacaaatgatalagtgtactcgttgcataaag 3881
 |||||
 DB 467 ATGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 526
 OY 3882 ttattggaatgaataaatccctctatattatagatgatatcgaactgcta 3941
 |||||

D	b	524	AAA	465
O	y	1332	gtgaaaaagaaagaaaagaacactccgcgaatgaaaacaggctgcctca	1397
D	b	464	AAAN	405
O	y	1332	ttaattgtgaatacatcatttcgaagcaaacaccgcttccagttagaacaatccc	1457
D	b	404	NNNNNNNAAAAAAAAAAAAAAAAAAAAAGCAWSTMAAAATTAATASTWMMWWTTT	345
O	y	1452	attctctaactgcctcttgaaactaatgttaaagttlaagttlaaacatalc	1517
D	b	344	GTTGTGWTWATTGGWRGTRMTTAAATTAMARRMAABRAWMAAGSTACRARAANAAM	285
O	y	1512	tagtgtatgatcatt	1525
D	b	284	AWAWAAWAAWATAGT	271

ACCESSION	AF068185	617 bp	DNA	GSS	07-MAR-2000
DEFINITION	tomato Lycopersicon esculentum genomic clone tox000011l6r, DNA sequence.				
VERSION	AF068185				
KEYWORDS	AF068185.1	GI:4221462			
SOURCE	tomato.				
ORGANISM	Lycopersicon esculentum				

REFERENCE	1 (bases 1 to 617)
AUTHORS	Budiman, M. A., Mao, L., Wood, T. C. and Wing, R. A.
TITLE	A deep-coverage tomo to BAC library and prospects toward development of an STC framework for genome sequencing
JOURNAL	Genome Res. 10 (1), 129-136 (2000)
MEDLINE	20113122
COMMENT	Contact: Wing RA

Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwinge@clemson.edu
Seq primer: GGAACAGCATGACCAACG
Class: BAC ends
High quality sequence stop: 476.

URES
source

```

1. 617
/organism="Lycopersicon esculentum"
/cultivar="Heinz 1706"
/db_xref="taxon:1081"
/clone="tox0001116r"
/clone_lib="CGI Tomato BAC library"
/tissue_type="Nuclei preparation from leaf"
/lab_host="E. coli DH10b"
/notes="vector: pGEM-BAC 11; Site.1: HindIII; Site.2:
HindIII; Tomato is a vegetable crop that ranks second only
to potatoes in value and importance. Among plant
geneticists and physiologists, tomato represents an ideal
dicot model beside Arabidopsis and monocot rice to derive
genomic information from. To facilitate the genome
analysis of tomato, we have constructed a tomato BAC
library that is suitable for positional cloning, physical
mapping, and genome sequencing. The library contains 129
,000 clones and a random sampling of 498 clones indicated
an average insert size of 117.5 kb. With 15X haploid
and genome equivalents (1C equals 953 Mb) (Armuganthan and
Bartle, 1991), the probability to recover any particular
sequence is greater than 99% High stability, large insert

```

and ease in manipulation make BAC libraries the choice for genome sequencing. Pre characterization of a few hundred bases of insert ends will make BAC clones extremely useful for rapid contig assembly (Venter, Smith, and Hood, 1996). Here we present the construction, characterization of the tomato BAC library, and preliminary analysis of the 1536 tomato BAC end sequences."

Query Match	1.2%;	Score 77.4;	DB 227;	length 617;
Best Local Similarity	47.0%;	Pred. No. 3.2e-06;		
Matches 274;	Conservative 0;	Mismatches 306;	Indels 3;	Gaps 1;

QY	4668	gaanaagaaataatgattatgacccctttagttagtcaaccgttatagacttg	4727
Db	613	GATATATACCAAGAGAGGGAATACATATATATGAGACTTTTGACACCGTTGGCAAGATG	554
QY	4728	tgaataagaattatgacataaataagtggtgaattaggaatgacaattcaacattg	4787
Db	553	AGGGCATATCAGAAATTTTATATACCTTTTCCTGCATTCATGAGGGGTTCAAGCTGTATCAAAATG	494
QY	4788	acgtcgagtcggtacatcaatgcctcattactcaatccaattatgcttc	4847
Db	493	ATGTGAACAGTGCATTTTCTAATATGAGATATCAAAAGAGGAGATTTGTGTCAACACCTG	434
QY	4848	ctcctaatactgtaacctttgaaagaaacacattggttgattgaaacgttcgtctatg	4907
Db	433	CTGGTTTTGAAGATGTAGAGCTACCAATCATCTGTTCACCTGTGAATTAAGCCTGTATG	374
QY	4908	ggttbaaacaagtcggttcttgtaacgtgatcaacactatcaaaagaglatgtaagacatg	4967
Db	373	GTCTGAACGAGCTCCCAAGACATGATGAATAAGCTGTCAAAATTTCTGCTGAATAATG	314
QY	4968	gtttactccaagttttaacaaatgagtgttattcattccattgatatgaagaggtcag	5027
Db	313	GTTTTCAAAAAGGCGCAGATAGACAAATCTTTGTTCTATTTPAAAAGAAACAAGAAATTGC	254
QY	5028	taatatatttgattatgattgatgatattcttacttgattgtaagttccaaagatta	5087
Db	253	TTTATCATTTCAAGTTT---ATGTGATGTACAAATTTTTGGAGCTACTTCAGAACATTTGG	197
QY	5088	ttagataatttggtgataccaattgagagatcatttgaagttaaagtggttggaatat	5147
Db	196	GGGAAGAAATTCATCATTAATAGCAAAAGAAATTTGAAATGAGATATGATGGTGAACCTGA	137
QY	5148	caaatctatctggtatgtaattcgtaaaacggaatctggttatatttactccaagaaa	5207
Db	136	CATTCTTCCTGGGTTTACAAATTCAAACGAATCATCAATGGGATTTTCAATCTGTCAGAGAG	77
QY	5208	aattctcaagaataacttaagaatttcaactagatgact	5250
Db	76	AGTACATTAAAGAGCTACTGAAGAAATTCATATATGTTTGACTC	34
RESULT 15			
A2548467		LOCUS	
DEFINITION		A2548467	908 bp DNA
ACCESSION		ENTREKJ0TR	Entamoeba histolytica Sheared DNA
VERSION		A2548467	genomic, DNA sequence.
KEYWORDS		A2548467.1	GI:11172102
SOURCE		SSS	
ORGANISM		Entamoeba histolytica.	
REFERENCE		Entamoeba histolytica	
AUTHORS		Eukaryota; Entamoebidae; Entamoeba.	
TITLE		1 (bases 1 to 908)	
		Loftus, B., Van Aken, S. and Fraser, C.	
		Determination of clone end sequences from Entamoeba histolytica	
		HMI:IMSS sheared DNA library	
		Unpublished (2000)	
COMMENT		Loftus	
		Contact: Brendan J	
		Loftus	
		14-NOV-2000	

Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: bjoftus@tigr.org

Clones are derived from the Entamoeba histolytica HMI:IMSS sheared
DNA library
Seq primer: M13-Reverse

Class: shotgun
High quality sequence start: 17
High quality sequence stop: 828.
Location/Qualifiers

FEATURES

source

1. .908
/organism="Entamoeba histolytica"
/strain="HMI:IMSS"
/db_xref="taxon:3759"
/clone_lib="Entamoeba histolytica sheared DNA"
/note="Vector: pHOSt1; Site: 1; Bst I; Constructed at The
Institute for Genomic Research (TIGR), Rockville, MD.
Genomic DNA isolated from broth cultures of E. histolytica
using a method described by Clark and Diamond (Clark,
C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
method for isolate identification. Exp. Parasitol.
77:450.). The DNA was mechanically sheared to give a
tight size distribution (~2 kb). The v + i method used for
the library construction is described in detail in Smith,
H.O. and Venter, J.C. (Making small insert libraries for
whole genome shotgun sequencing projects. In Genome
Sequencing: A Practical Approach, eds. M. Vaudin and B.
Barelli, Oxford University Press, 1999)."

BASE COUNT 434 a 16 c 289 g 169 t
ORIGIN

Query Match 1.28; Score 75.4; DB 245; Length 908;
Best Local Similarity 44.18; Pred. No. 9.3e-06;
Matches 316; Conservative 0; Mismatches 401; Indels 0; Gaps 0;

```
Oy 3588 aagaacagctccagaagatgctactctgactccagaagaactagtcgaagatg 3647
Db 23 AAGAAGATGATGAGATGATGATGATGATGATGATGATGATGATGATGATGATG 82
Oy 3648 aacagcaaatccctcgtactgctaaataagaaatcgggtactgtaaaataagatgag 3707
Db 83 AAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 142
Oy 3708 gagaataattcattccgggggggtgatgatgattcgtcgatcaactcaaatgctg 3767
Db 143 ATGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 202
Oy 3768 agcaatcattatgtaaaacagagatgctgtaacagtcacatccaaatcagaagttctc 3827
Db 203 AAGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 262
Oy 3828 aagaaggagaataacttaataaactgataatgtaactgctgtaaaatgtaagttatg 3887
Db 263 ATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 322
Oy 3888 agaatgaaaaaatctcctcatttaattcaatagatgataactgtaactgctacagact 3947
Db 323 ATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 382
Oy 3948 cgggaatgatgcaatccaaatccgaatcccaatccgaatccgaatccgaatccgaatcc 4007
Db 383 AAGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 442
Oy 4008 tgattaatgagaataatactgaaataatccaaataacacatgaaagtaatccctgctgata 4067
Db 443 ATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 502
Oy 4068 agagattggaatgtaaacgtataatgttgatgaaatggaatggaatggaatggaatg 4127
```

```
Db 503 AACAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 562
Oy 4128 atgacattgctgaagctaaatccactcagatgaaataatgattctcgaatgaga 4187
Db 563 ATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 622
Oy 4188 gtttgataataatcattgacatgctcagagcaagaagaataacacattgagaagaag 4247
Db 623 ATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 682
Oy 4248 ttaacgaaaaaatgctggtgacttaaacattcactgatacactgataccaagaagaag 4304
Db 683 ATGAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 739
```

Search completed: August 25, 2001, 00:08:01
Job time: 6426 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 24, 2001, 22:49:15 ; Search time 8259.91 Seconds
(without alignments)
12033.516 Million cell updates/sec

Title: us-09-430-590E-3

Perfect score: 6426
Sequence: 1 tcttggttctgcacattt.....agaagttatattccatca 6426

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1344157 segs, 7733874588 residues

Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl:*
1: gb_ba1:*
2: gb_ba2:*
3: gb_ba3:*
4: gb_in1:*
5: gb_in2:*
6: gb_in3:*
7: gb_ov:*
8: gb_ov:*
9: gb_pac1:*
10: gb_pat2:*
11: gb_ph:*
12: gb_pl1:*
13: gb_pl2:*
14: gb_pl3:*
15: gb_pl4:*
16: gb_pl4:*
17: em_ba2:*
18: em_fun:*
19: em_htgo_hum:*
20: em_htgo_hum:*
21: em_htgo_hum:*
22: em_htg_hum1:*
23: em_htg_hum2:*
24: em_htg_hum3:*
25: em_htg_hum4:*
26: em_htg_hum5:*
27: em_htg_hum6:*
28: em_htg_hum7:*
29: em_htg_hum8:*
30: em_htg_inv1:*
31: em_htg_inv2:*
32: em_htg_other:*
33: em_htg_rtd:*
34: em_hum1:*
35: em_hum2:*
36: em_hum3:*
37: em_hum4:*
38: em_hum5:*
39: em_hum6:*
40: em_hum7:*
41: em_in:*
42: em_ov:*
43: em_ov:*

44: em_ov:*
45: em_pac:*
46: em_ph:*
47: em_pl:*
48: em_ro:*
49: em_sts:*
50: em_sy:*
51: em_un:*
52: em_vl:*
53: gb_sts1:*
54: gb_sts2:*
55: gb_sts3:*
56: gb_sy:*
57: gb_un:*
58: gb_vl1:*
59: gb_vl2:*
60: gb_htg1:*
61: gb_htg2:*
62: gb_htg3:*
63: gb_htg4:*
64: gb_htg5:*
65: gb_htg6:*
66: gb_htg7:*
67: gb_htg8:*
68: gb_htg9:*
69: gb_htg10:*
70: gb_htg11:*
71: gb_htg12:*
72: gb_htg13:*
73: gb_htg14:*
74: gb_htg15:*
75: gb_htg16:*
76: gb_htg17:*
77: gb_htg18:*
78: gb_htg19:*
79: gb_htg20:*
80: gb_htg21:*
81: gb_htg22:*
82: gb_htg23:*
83: gb_htg24:*
84: gb_htg25:*
85: gb_pr1:*
86: gb_pr2:*
87: gb_pr3:*
88: gb_pr4:*
89: gb_pr5:*
90: gb_pr6:*
91: gb_pr7:*
92: gb_pr8:*
93: gb_pr9:*
94: gb_ro1:*
95: gb_ro2:*
96: gb_in4:*
97: gb_pr10:*
98: em_ba3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6426	100.0	6426	12 AF007776	AF007776 Candida a
2	6421.2	99.9	6980	12 AF050215	AF050215 Candida a
3	432	6.7	4872	14 CAL251464	AJ251464 Candida a
4	400	6.2	408	12 AF030556	AF030556 Candida a
5	339.2	5.3	1470	12 AF078809	AF078809 Candida a
6	174.6	2.7	1017	53 CNS078UD	AL437951 T7 end of
7	166	2.6	1888	10 AX073190	AX073190 Sequence
8	142.4	2.2	905	53 CNS07ACX	AL436423 T3 end of

9	127.2	2.0	838	53	CNS07ARP	AL436955 T7 end of
C	10	99.4	1.5	80561	12 AC073555	AC073555 Arabidops
C	11	98.8	1.5	109155	14 ATT20K12	AL137898 Arabidops
C	12	98	1.5	85961	12 AC079131	AC079131 Arabidops
C	13	98	1.5	103637	12 AC079604	AC079604 Arabidops
C	14	97.2	1.5	97146	12 AC007259	AC007259 Arabidops
C	15	96.4	1.5	4793	13 ATLECR6K	Y08010 A.thaliana
C	16	96.4	1.5	13558	14 ATT93161	U93161 Arabidopsis
C	17	96.4	1.5	91851	14 ATT16L24	AL138659 Arabidops
C	18	96.2	1.5	1035	53 CNS07AOP	AL436847 T7 end of
C	19	94.6	1.5	82612	14 ATT31B5	AL163491 Arabidops
C	20	93.2	1.5	82875	12 AC013482	AC013482 Genomic s
C	21	92	1.4	175565	13 AP003311	AP003311 Oryza sat
C	22	91.4	1.4	16720	14 LEU68072	U68072 Lycopersico
C	23	89.8	1.4	5846	14 PVA005762	AP005762 Phaseolus
C	24	89.2	1.4	111554	13 AP002460	AP002460 Arabidops
C	25	89	1.4	152423	12 AC079853	AC079853 Oryza sat
C	26	89	1.4	192174	83 AP003263	AP003263 Oryza sat
C	27	88.8	1.4	145789	60 AC007926	AC007926 Trypanoso
C	28	86.2	1.3	84785	13 AP002064	AP002064 Arabidops
C	29	86	1.3	157788	83 AP003289	AP003289 Oryza sat
C	30	85.6	1.3	154662	72 AC051632	AC051632 Oryza sat
C	31	85.4	1.3	72520	12 AC006570	AC006570 Arabidops
C	32	85.4	1.3	77290	12 AC0074284	AC0074284 Arabidops
C	33	84.4	1.3	123183	12 AC006841	AC006841 Arabidops
C	34	84	1.3	110000	84 PFMA14P1_1	Continuation (2 of
C	35	84	1.3	144322	13 AP002867	AP002867 Oryza sat
C	36	84	1.3	145358	78 AC091086	AC091086 Oryza sat
C	37	84	1.3	147857	13 AP002747	AP002747 Oryza sat
C	38	84	1.3	153477	60 AC006278	AC006278 Plasmodiu
C	39	83.8	1.3	13341	14 OLKRLP1	Y27126 Oryza longi
C	40	83.4	1.3	9093	14 AY016208	AY016208 Arabidops
C	41	83.4	1.3	87286	12 AB026651	AB026651 Arabidops
C	42	83	1.3	151978	13 AP002069	AP002069 Oryza sat
C	43	82.6	1.3	153776	83 AP003511	AP003511 Oryza sat
C	44	80.8	1.3	155215	83 AP003449	AP003449 Oryza sat
C	45	80	1.2	189349	12 AC066654	AC066654 Genomic S

ALIGNMENTS

RESULT 1
AF007776 6426 bp DNA PLN 21-NOV-1997
DEFINITION Candida albicans retrotransposon pCal, complete sequence.
ACCESSION AF007776
VERSION AF007776.1 GI:2636718
KEYWORDS

SOURCE

ORGANISM Candida albicans.
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; mitosporic Saccharomycetales; Candida.

REFERENCE

AUTHORS Matthews G.D., Goodwin, T.J., Butler, M.I., Berryman, T.A. and
Poulter, R.T.
TITLE pCal, a highly unusual Ty1/copia retrotransposon from the
pathogenic yeast Candida albicans
J. Bacteriol. 179 (22), 7118-7128 (1997)

MEDLINE

AUTHORS Matthews, G.D., Goodwin, T.J.D., Butler, M.I., Berryman, T.A. and
Poulter, R.T.M.
TITLE Direct Submission
JOURNAL Submitted (11-JUN-1997) Department of Biochemistry, University of
Otago, Box 56, Dunedin, New Zealand

FEATURES

Source Location/Qualifiers
1..6426
/organism="Candida albicans"
/db_xref="taxon:5476"
/transposon="retrotransposon pCal"
1..280
primer_bind 281..291

/note="Minus-strand primer-binding site"
398..1372
/note="ORF1"
/transl_start=1
/transl_table=12
/product="GAG protein"
/protein_id="AAC49879.1"
/db_xref="GI:2636720"
/translation="MSSAKNDNEGVESVDQANASVDEHAKAFNMLEIFKPNL
PKLVGNOKSVDMKNEEFYFVAAPDYLEFLDYNDPKDKVKKVEGYFETGCMLO
CIOSIFDRRLMISKLPKHLKOKKANLKAADATYKSDYITTSKIIISKVYNEHEL
VVCYNLPYLSQVEKELETLNFTSVNVEVRSPLNLIGVYFNHVKSEALSLFLN
IHASYSKMLQADNDTSVLPSCSTIAEEMCDPDAKRLVDIPSNKEELNLVSLPAP
KPKGKEENSOSOKNKKSRKRNKKHPSKSDNGEKEKEKTSSE"
1370..1372
/note="potentially suppressed stop codon"
41373..6103
/note="ORF2"
/codon_start=1
/transl_table=12
/product="POL protein"
/protein_id="AAC49878.1"
/db_xref="GI:2636719"
/translation="KTGAASINCVANINHCSTTPPVNSHSLASLVANMFKLRN
KLYVDGTATISVNNKDLISVVKDATTIEVSADATLEADIGDLIRVAVISTTE
NPLYPSESEFNLVSLKQIEERGFNLKESYIVNOMVAPPIIASRNADLVWGP
FSESLCEPDFDGLADMLSNODKSESNMESSEVOEHYSRALINSLTEVDV
DVEISRYGVEOLPTGDKNDIYFNHLSHMSIEKILLKQGLVLTSTRESLOKIA
DKKCLISNAKORSHNHSEKRSRHEHEDTIGPSESNKKVYLTVSVIDEHTGY
EGITIKDRKVRKLLIQRLLKINNRNDKAYATRSNMAEFPPPSLAFFGIMRETTA
YSPELNGLEAVNKLILDOITRTIVTFSLGQILKLIYVYQISYIMINHTPRSLKGT
PYGCYQYLSSEGNFRPEPPIIDCVTFESNAIEKNRKYSTGAPSSINGAVIGASDC
FSYVYLKMRCDIILSPVRLIRSEYVINSYKMLSTPMSHIVPMAEGIGRSGA
OYEVRTYVESEYDNTNDVMHMKSEYVOPASFLTGNSENEVINDDPVQITEN
PDPSNPILQITRESDMVESEVSDNPMPSLHETLPGNPPVSKPPOLETETSGVSKS
EPTNHTKAPISIGRDRKRSSTVQVLSHPQCTGTAESKISGTDHRCGVVKE
TVSEDMHTSDIPETSAEDQNPVLSANKNRTEKIDGENTISFPGGDSDVYNSV
EOSNVEDAGNSPIODEVSGRLINTEQTLIDVAVAIENEKISPIINSIDHTELA
TDSGNSNSTESDIOSKNEISPIVINEKMTIELIOKHIESILADKRLDEFTYNDIEIN
VINDDIAEANLPDENNDVONNESSESRKAKRYFEKEVNEKIACTGKSLHD
TDPREAIRVLTNGETKRTEPKREVPIYVYKLSKRSOYSPVYRGRGIVAPKRYLA
VYKTIYDNDPGWIKSMNALEKFRSKDYEVPIPTGVKPIISMGMVHEKIDSLKGV
KRSQVYHNRKREKLDIDPEFSVSSVIDVLTIRLLITIIIGELGHTIOHLDVESYILN
ASITSHNPIYFPEKSVPLKKNHMLKRSYVGLKOSGLEWHHTKRLVLEDIGETVYL
HNDGFLHEIEEGSVIYGLVYDDILMWSQKVDNEDVDLRHFEKVGELISNYL
GIEPRTESGYILISOEFLKLLKDFKLDSSGKNIPWIPMDKEVAKIIEENPEN
DEKVPNERTLDDPAKLLYOSGVSLMAATNTRPDISVYVNSGSKCANPNVDYEC
LTYCLRYIKNSGCHIEYKRNMLNTPKSPYIEGCSDFAPGSDRKSISTILYVNC
NLQWATKQIVIAOSSACEMALNLTMLKAIETKNHMDLGEVGIHODNOAY
IKVLNRNCHPRPIDIQYKFLRQLINKVFISVYKTDNVDACMTCLSRNFKAF
VEGMIRKLDLEPNOTSIGNAIVAE"

misc_feature
1373..1380
/note="Region: putine-rich tract"
misc_feature
1381..1455
/note="Region: potential pseudoknot"
misc_feature
3455..3465
/note="Region: polypurine tract 2"
misc_feature
6136..6146
/note="Region: polypurine tract 1"
LTR
6147..6426
BASE COUNT 2228 a 1031 c 1259 g 1908 t
ORIGIN

Query Match 100.0% Score 6426: DB 12: Length 6426:
Best Local Similarity 100.0%: Pred No. 0:
Matches 6426: Conservative 0: Mismatches 0: Indels 0: Gaps 0:
QY 1 tgttggttgcactatttgcgaagaacgcatcaatgaagaatgagtattatga 60
|||||
Db 1 TGTGTGTTTGCACATATTGTGTCAGAAACGATCATTGAAATGATGTTATTATGA 60
QY 61 gaatgaagaattttccatcaacatcagtgatgacagaactaaactatattgtgtagt 120

Db 61 GAATGAAAAATTTTCCATCCACATCAGTGTGATGACAACTTAACATATATGTGTAGT 120
Qy 121 ataaataaggtatgaataatcaccacatccccaataatacaagagatagaagggagagt 180
Db 121 ATAAATAGGAGTATGAATATCCACATCCCAATATATCAGAGATAGAGGAGAGAGT 180
Qy 181 tcaatatatacttctgtaataaactcgtcttaattcaactatacacactagaagct 240
Db 181 TCAATATATATCTGTGATATATTAAGTCTGTCTTAATTCATATACACACAGTACGAGT 240
Qy 241 aacacctcaatctcaggtgaagaagaagttatattccatcagaattgaagtcgatagtat 300
Db 241 AACACGCTCATCTCAGGTAAAGAAAGTTATATTCATCAGATGATGAAAGTCAATAGTAT 300
Qy 301 aatcaattcgtcccaaatagcgttctgataaactcagtcaccagaatttgatattatgatt 360
Db 301 AATCATTTTGTCTCCCAATATAGGTTGTATTAATTCAGTCCCTCAGATTTGTATTTGATTT 360
Qy 361 gataagttcgaagtttgaagtlacagaattcacaagaatgattccgcaagaatgatga 420
Db 361 GATAGTTTCGAGGTTTGAAGGTACAGAAATTTCCAAAGATGAGTCCCAAGAAATGATGA 420
Qy 421 taacgaaggagagtcatacgaagaagtggttgaatcaagaatgattatagtaagtgatga 480
Db 421 TAACGAGGAGGAGTCATGGAAGGTGATCAAGCTTAATGCTAATGCTAATGATGAAGTGA 480
Qy 481 acatacgaagtcataatcaatcgttctcaataaattgaacttaacaaagtlagc 540
Db 481 ACATATCAAGGCTAGATTCATATATGCTTTTCATTAATTAATTAATGACTTACCTAACTT 540
Qy 541 cgtcgtgaatcagaagaagcgttgataaattgaaatgaagaatttaataattccacgttgc 600
Db 541 CGTGCTATATCAGAAAGGCTGATGAATGAATGAAGAATTAATTAATTTCCACGTTGC 600
Qy 601 ttaccccgatgttttgaatttttgcctgacataatacttaagaattcaaggttaa 660
Db 601 TTACCCCGATGTTTGGAAATTTTCTGACTAATATCCTTAAGAAATTAATTCAGGTTAA 660
Qy 661 aaaggtaagaagttatatttactcgtgtgtgttacaagaatgttttcaagtcattt 720
Db 661 AAAGGTAAGAGTATTTATTTTACGTGTTGCTTTACAAATGCTTTTACACTCATTTT 720
Qy 721 tgataggttcagatgcatacgaattcttaagctacaaagcattgcgaagaagcaaa 780
Db 721 TGATAGGTTCAATGATGATCATGATTTCTAAGCTACAAAGCACTTGCAAAAGGAAGCAA 780
Qy 781 cctaatcaaaagctgtcttgaatgctgttacttaactaactaagaattataccattacgtaa 840
Db 781 CTTAATCAAAAGCTGCTTATGATGCTGTTACTAATCTAAGATTAATACCATTTACTAGTAA 840
Qy 841 gatcttgcgaagtttctgaacgcttgaacatgagttagtggttctgaataacattcaca 900
Db 841 GATCTGCTGAGAGTTTGTAAACGTTGAACATGAGTTAGGTTGCTTAACCTTCCATA 900
Qy 901 ttgtctgcaggtggaagaagaacttgaggaataactcacaacactcaaacgtttgtcga 960
Db 901 TTTGCTGAGGAGTGAAGAACTTGAGGAATACTCTTACAAACACTTCAAAAGCTTGTGGA 960
Qy 961 tgaagtgtccgtagtcttccaactcctatagtgcaagttcttgtaactcaactatgtaa 1020
Db 961 TGAGTATGTCCTAGTCTTCCAAATCTATAGTCAAGTCTTGTGATCTTCAATCATGTGAA 1020
Qy 1021 gaaatcagaagcctttaaagtttcttgaatatcatalgctcctatactactacaagtgat 1080
Db 1021 GAAATCAGAGGCTTAAAGTTTGTGATATTCATGCTCCATCATCTCAAGTGTAT 1080
Qy 1081 tcaagctgacaatgataatcagtaactcccaagtgctctacacatagctgaagaatgtg 1140
Db 1081 TCAAGCTGCATATGATATCATGATCTCCAAAGTTGCTCTCCATAGCTGAAGAAATGTG 1140
Qy 1141 tgatcacctgtatgtctagaattggttgacatcccaagcaacaatatgtaacttaact 1200
Db 1141 TGATCATCTGATTTATGCTAGTATGTTGATTCACATTCACAGCAACAATATGACTTAATCT 1200
Qy 1201 tatgttagtttaccacacccagagaacccaagaagaacccagagagatcactcgtga 1260
Db 1201 TATGTAGTATGTTTACCACACCCAGAGAAACCAAAAGGAAAAACAGAGGAGATCAGTCA 1260
Qy 1261 acaatcccaaaagaagaacctgaaatcaagaagaagaataagaagaatcccaaatcaga 1320
Db 1261 ACAATCTCAAAAGAGAACCTGAATATCAGAGAAAGAAATTAAGAAATCTCAAAATCAGA 1320
Qy 1321 taacgaataaagtggaagaagaagaagaagaagaacttcactcgtgaatgaaagaag 1380
Db 1321 TAACGATTAAGCTCAAAAGAAAAAGAAAAAGAAAAAAGTCACTGGAATGAAGAAACAG 1380
Qy 1381 tgcgtctctatcaatgtgttaagaatatataatctgaagaagaagaacccgttccagt 1440
Db 1381 TGCTGCTCTTATTTATTTGTGATGAATATATCATTAATGAGCAAAACCAAGTTCCAGT 1440
Qy 1441 agaaatctcactcctccttaatgcttcttgaacgttaatgaaatcttaaggttgaagt 1500
Db 1441 AGAAATTTCTCATTTCTTAAATGCTTTTGAACGTAAATGAATTTAAAGTTTAAAGCTT 1500
Qy 1501 taacaagatcagtgatgatactatggtgccacaatactcgttctgtaacaataagaat 1560
Db 1501 TAACAAGATCTAGTATGATGATGATGCTGACCAATATCTGTTGTGMAACAATAAGATAT 1560
Qy 1561 atgtcgaatgttaagagccaacaatggaagttcgttgcctgagtggtcactataga 1620
Db 1561 ATTCTGATATTTAAGAGCCCAACAATTTGAATTTCTGTTGCTGATGAGTCTCATTTAGA 1620
Qy 1621 agcgaattgtatgtgatccttaattacagaagtcgtatgtctcgaattcagttagaag 1680
Db 1621 AGCAGATTTATGATGATGATTAATTAATCAGAGTCGATTTGCTCCATTTAGGTTAGGA 1680
Qy 1681 tactgtatcttaccagaagaagtccttctaactctgtgaagtttgaacaattgaagaag 1740
Db 1681 TACATTTATTTTACCAAAAGTTCTTTAATCTTGTGAGTTTGAACAAATTAAGAACG 1740
Qy 1741 aggaattgaatgtcttactaagaagaacagagatgttcttaaccaaaatgtgctcc 1800
Db 1741 AGGATTTAATGCTTTTATTTACTAAGAAATCAGTATGTTTAAACCAAAATGTGCTCC 1800
Qy 1801 tactattatgtctcaagaagaagatgtcgtgaatcttataatgtgtcctcaatcagtg 1860
Db 1801 TACTATTTATGCTTCAAGGAAGAAATGCTGCTGATTTATATGAGGCTCAATTCAGTGA 1860
Qy 1861 agaatcttgaagaatgtattgattatgattatgattgttggcagatatgttgcgaatgctaa 1920
Db 1861 AGAATCTTTAGAAATGATTTTGTATTAATGATGATTTGGCAGATATGTTGTCCAATGCTAA 1920
Qy 1921 ccaagatgacaagaataatcaagtaatgaatgtcagaatatcaagaacatgatta 1980
Db 1921 CCAAGATGACAAAGATTAATCAAGATGATGAATGCAATTCGAATATCAAGAAATCATTA 1980
Qy 1981 taagttctcgagcattaaatactcttgacgagagttgattgtttagatgttgaatttc 2040
Db 1981 TAGTTCTCGAGCAATTAATAATTTCTTGAAGGAGGTTGATGTTTGAATGTTGAATTC 2040
Qy 2041 cccatgtgaggttgaacaacttgcctacaaactggaagataagaagatatattatattcca 2100
Db 2041 CCCATATGAGAGTGAACAATTTGCTACCAACTGAGATGAAGAAAGATTAATTAATTTCCA 2100
Qy 2101 ttgatgtcaaatcatalgtgccattggaataacttctgttctcaaaaaatacagaaggtc 2160
Db 2101 TTGATGTCAAATCATATGTCATTTAGGAATAATCTTGTCTTACAAAATATCAGAGGTCT 2160
Qy 2161 cgttaactcacaactcaaaagaagtcctcaaaagattcgtatgttgaagttatgtcatt 2220
Db 2161 CGTACTTCACACTTCAAAAGAGAGTCTTCAAAAGATTTGCTGATTTGAAGTATGCTATT 2220
Qy 2221 atcgaatgcacaagaagaagtcacaatactcaatcaagaagaagaagcctcgaagaagaa 2280
Db 2221 ATCGAATGCCAAACAGAGAGTCAACATCATTCATTCAGAAAGAAAGCCTCGAAGAGACA 2280

QY 2281 tgaagagcttcatgtatatactctcgtccattagctccgaaataaagaatggtatt 2340
DB 2281 TGAGAGCTTCTATTTGATATCTCGGTCCATTAGTCCGAAATATACAGTGGTATTT 2340
QY 2341 aagctcgttataagatgaataacaggggttaccatgaagaattatataaagaacagaa 2400
DB 2341 AAGCTGTTATAGATGACATACGGGTACATTTGAAGAAATTTACTTAAGACAGAAA 2400
QY 2401 ggttaagaatcctcttaattcaagaataaagatcctggaataatcggtttaagaataagt 2460
DB 2401 GGTAAAGATCTCTTAATTAAGCATTAAGATCTGGAATTAATCGTTTAACGATAAGGT 2460
QY 2461 ggcatactcaagaatataatgtcctcgtgagttcccaacaccttgcattagctgaagt 2520
DB 2461 GGCATCTTCAAGAGTATATGCTCTGAGTCCCAACCTTCCTGATTTAGCTGAGTT 2520
QY 2521 cggatattggaaggagatacagcgacatacctgctgagcttaagtctcgcgaggt 2580
DB 2521 CGGATTTGGAGGAGACTATAGCGCATATCTGCTGAGCTTAATGCTCCCGAGGT 2580
QY 2581 tgttaataaattgatttacaacagatttacagatcgttggtaagacttggccacaagt 2640
DB 2581 TGTTAATAATTTGATTTTACACAGATTTACAGATGCTTGTGACACTTGTGTCACAAAT 2640
QY 2641 actcaagttgattatattatgtatcaataatctcattacaatgatcaacacacacacacg 2700
DB 2641 ACTCAAGTTGATTTATGATGATTTCAATTTCTATTACAAATGATCAACACACTCCAG 2700
QY 2701 tctgttaactcaaggagaacacacctatggttgcattatcaataaagtgaaggaaatt 2760
DB 2701 TCGTTTCACTCAAGGACAAACCCCTTATGATGCTGATTAATCAATTAAGTAGGAGAAATTT 2760
QY 2761 ctaccggttcccttgcctcagatgtgtcgttaccattagtaatccatcgaaagaa 2820
DB 2761 CTACCGGTTCTTTCGCTATGATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2820
QY 2821 ccgttaccgagttacatacactaaaggagcctccatcagatcagtggtcgttgat 2880
DB 2821 CCGTTACGAGATTCATCACTAAAGAGGCTCTTCATGATCATGGGTCTGATGATG 2880
QY 2881 ctacgagtgagattggtttagtattacggttgccttcaaaaataatgagtgatattat 2940
DB 2881 CTACGAGTACGATTTGATTTATGATTTATGATTTGCTGCTGCTGCTGCTGCTGCTGCTG 2940
QY 2941 ccttagccccaatgtccgatatatgcgaagctatgaggttattcaaccctatccaaa 3000
DB 2941 CCTTAGCCCTATGCTCGATATGCGAGCTATGAGCTTATTAACCTCATCTCAAAA 3000
QY 3001 ctatcaactaacctatgtcaacatgttctcctatgtcgtgaaggtatccagggaagca 3060
DB 3001 CTTATCCACTACCTATGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3060
QY 3061 acttgagagctcaag 3120
DB 3061 ACTGAGGCTGATGATGAG 3120
QY 3121 tgaagtgatgacatgcccgaagatcatattcagttcagccagacatcgttacttaac 3180
DB 3121 TGAGTGATGACATGCCCAGAGATCATATTCAGTTCAGCCACATCGTTTACTTTAAC 3180
QY 3181 taaggttaacagttcctaagaatatgttataatgatgatccagttacagatccatctga 3240
DB 3181 TACGGGTAAAGCTTCTAACGAAATGTTATTAATGATGATGATGATGATGATGATGATG 3240
QY 3241 gaatcccgatgatttttctaaacctctcaactaactgaagaatcaacagatatgtatc 3300
DB 3241 GAATCCCGATGATTTTCTAAACCTCTTCAACTAAGAAATCAACAGATATGATATC 3300
QY 3301 cgaagttaaatcgaatgaatacttaaacccagctctccagagcttaacacctgggagtaa 3360
DB 3301 CGAAGTTAAATCGATGAGATCTTAACCCAGTCTCCAGAGCTTAACACCTGGGAGTAA 3360
QY 3361 tccggtgtctaaacctccctcaacttgytaccgagactcaagtaataaggagttcaaga 3420
DB 3361 TCCGGTGTCTAAACCTCTCAACTTGTGATCCAGAGCTCATTAATAGGAGTCTTAAGA 3420
QY 3421 gccattcaacaacacacaaaggagcgccttccatccaggaggagagacataagcct 3480
DB 3421 GCCATTCAAAACACACAAAGAGAGGCCCTTCCATCCAGGAGGAGGCCATTAACGCT 3480
QY 3481 ggaatctactcgtcaggttggacttatacacaaccccaagcttactccgctcgga 3540
DB 3481 GGAATCTACTGCTCAGGTGGAGTATCAACCCACCCAGCTGATCTCCGCTTGGA 3540
QY 3541 ggaagtaaatgttcaagaagacatcattcgtgtcagctgtttaaagaagaagctc 3600
DB 3541 GGAGTCAAAATGTGAGGAGAGATCATTTGGTGTCCAGCTTGTTAAGAAACAGTCTC 3600
QY 3601 agaagattggcactctctgactaccagaacacagtgctgaagaatgaacagcaaatcc 3660
DB 3601 AGAAGATTGGCTACTCTTGTGATCCAGAAATGATGCTGAAGATGAACAGCAAAATCC 3660
QY 3661 ctcgttactggtcaataagaatcggtaactgaaanaaataagatgagagagaatatcc 3720
DB 3661 CTCGTTACTGCTATTAAGAAATCGGCTAACGAAATAATATGATGAGGAGAAATATTTC 3720
QY 3721 attccggggggtgagtgatctgtcgtgatacactcaaatglttgagcaatcaatgt 3780
DB 3721 ATTTCCGGGGGTGATGATGATTTGCTGCTGATCACTCAATATGTCAGATTAATGT 3780
QY 3781 tgaacagagagtgctgttgaagctcaatcaagaagaatcttccaaaggagaagaat 3840
DB 3781 TGAACAGAGAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATG 3840
QY 3841 acttaatgaacaaactgatalatagttgatalactgttgciaaagltatlgagaatgaanaat 3900
DB 3841 ACTTAATGAACAACATGATATGATGATGATGATGATGATGATGATGATGATGATG 3900
QY 3901 cctcctatlaattcaatagatgatacgaactcgtctacagactcggggaatgtag 3960
DB 3901 CCTCCTATTAATTCAATAGATGATGATGATGATGATGATGATGATGATGATGATG 3960
QY 3961 caatcaacagatccagatcaatcaatcaagaatgaatgaatgaatgaatgaatgaat 4020
DB 3961 CAATTCACAGATCCGATTCATTCGAAATGAATATGATGATGATGATGATGATGATG 4020
QY 4021 aatatcgaataatccaaaacacacatgaagaatcctcgtgataagagatlgatga 4080
DB 4021 AATATGTAATATATCAAAACACATGGAATGATGATGATGATGATGATGATGATG 4080
QY 4081 attgaaagctataatgttgaagaatgagagatgtgattatgacgtatgactgtga 4140
DB 4081 ATTTGAAAGGTATATGATGATGATGATGATGATGATGATGATGATGATGATG 4140
QY 4141 agctaatccactcagagatgaataatgaatgttcaagtgaatgaatgttgaataa 4200
DB 4141 AGCTAATCCACTACAGATGAATATATGATGATGATGATGATGATGATGATGATGATG 4200
QY 4201 tcatagcatgtcagagcaagaagaatatacatttgaagaagaatgaagaagaat 4260
DB 4201 TCATAGCATGTGACGAGCAAGAAAGAAATACATTTGAGAAAGAAATGAGTAAAG 4260
QY 4261 tgcgttactaaacattcaactgttatacaactgaatccaaagagaagaatcagaagttaa 4320
DB 4261 TGCCTGTACTAAACATTTCACTTGTATCAACTGATCAAGAGAGAAACCAAGTGTAA 4320
QY 4321 tactgtgaacccaagagatcgaacccaagaagaagaagtggtcctatccgtgtgaat 4380
DB 4321 TACTGTGAACCAAGAGATGAACCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4380
QY 4381 aaacaaagatcgcaataaagtcacatatttcaagaagtggtgaagcgttataa 4440
DB 4381 AAACAAAGATCGCAATATCAAGTCAACATATGTCAGAGAGAGAGAGAGAGAGAG 4440
QY 4441 cccaagaggtatttataatgtcgtcgaacaaatcgaataatgaatcggagtgat 4500

Db 4441 CCCCAAGGATTTACATCGCGTCGCACAAATGCACTATATGATCCGGATGAT 4500
 Qy 4501 aaagtcataatgctgaactagaagaatttaatacaagaagtttaagaagaattcc 4560
 Db 4501 AAAGTCATAAATGCTGAAGCTAGAGAAATTTAGATCAAAAGATTTAGCAAAAGTTCC 4560
 Qy 4561 aattccacacggtgctgaagccialatclatgggttggtatcacactgagaagaattgatic 4620
 Db 4561 AATTCCACCGGTCGTGAAGCTATATCTATGCGTTGGGTGATACAGAGAAATTTATTC 4620
 Qy 4621 tctcaaaagtgctgcggaatacagcttggttgctacatgcaagaagaagaaga 4680
 Db 4621 TCTCAAAAGGCTGTCGGAATCAAGTTGTTGTCATGCAACAGACAGCAAAAGGAAA 4680
 Qy 4681 attggattatgaccttggttggttgcacactgtatagacttggtgactataagatt 4740
 Db 4681 ATTGGATTATGACCCCTTTATGTTAGTTGTTACCTGTTATAGATCTTGACCTATAAGATT 4740
 Qy 4741 attgacaataataggttggtgaattagaatgacaattcaacattagaagctcgaagtcgcg 4800
 Db 4741 ATTGACAATATATGTTGTTGTAATTAGAATGACAAATTCACATTTAGAGCTGACGCGC 4800
 Qy 4801 gttactaatgctcttactatcattcaatccaaattatgcttctcccttaaatcaag 4860
 Db 4801 GTATCTAAATGCTCTATATCTATTCATTCAAATCCAAATTTATGTTCTCTTAATACGT 4860
 Qy 4861 accttgaagaagaacacatggttggttattgaagcttgcctatgaggtttaaagaagtc 4920
 Db 4861 ACCTTTGAGAAACCAATGTTGTTGTTATGTAAGCTGCTGCTATAGGTTAAACAGTC 4920
 Qy 4921 gggcttggaatggtatcacacactcaaaagagatggaagacatggttactcaag 4980
 Db 4921 GGGTTGGAAATGATCACACCTATCAAAAGATATTGGAAGATTTGTTTACCTCAAT 4980
 Qy 4981 tttaacaatagatggtttatttccacttgaatagaagggagtcagtaataattag 5040
 Db 4981 TTTAACAATGATGTTTATTTTACATTAATGTAAGAGGCTGATATATTTAAG 5040
 Qy 5041 tttaatagttgaatcattcttattggttggaagttcacaaaagattgtaaatgtt 5100
 Db 5041 TTTATATGTTGATGATATCTTATGTTGTAAGTTGCAAAAAGTTATGATATTTTGT 5100
 Qy 5101 ggaatcaatgagagatcatttgaagttaaagtggttggtgaataatcaaatatctg 5160
 Db 5101 GGATCAATGAGAGATCATATTTTGAAGTTGTTGTTGTAAGTTCAATATTTATCTGG 5160
 Qy 5161 tatgaatttcgtaaaacggaatctggttatatttatactcaagaagaatttctcaaga 5220
 Db 5161 TATTGAATTTGTAAGACGAATCTGTTATATTTATCTCAAGAAATTTCTCAAGAA 5220
 Qy 5221 attacttaagatltcaaaactagatgactcataatggaagaacataccctgattccga 5280
 Db 5221 ATTACTTAAGGATTTCAAACTAGATGACTCATATGSGAAAAACATTAACCTGATTCGAA 5280
 Qy 5281 tgacaaatcagaaaggttgcaataatctcgtaaaaaacgttaatccagagatgatttga 5340
 Db 5281 TGACAAATATGAAAAAGTTGCAATATATGCTGAAAAAGTTTATCCAGAGATGATTTGA 5340
 Qy 5341 aaaggttcgaatgagacttgcttgacccctgactataaaaaactataccaagaatggtgt 5400
 Db 5341 AAAGGTTCCGAATGAGACATGCTTGACCTGATGCTAAAAAATATACCAAGAGGTTGT 5400
 Qy 5401 tggcgtgcttattaggtcgcgcaacaacacgctccagatatacgctgctgtagtaattc 5460
 Db 5401 TGGCGTCGCTTTATGSGGCTGCCAACAACACAGCTCCAGATATATCGCTGATGTAATTC 5460
 Qy 5461 gttgggttctaaatcgcgaatccaaatgctcagatattagaatgaaattgattatgct 5520
 Db 5461 GTTGGGTTCTAAATCGCAAAATCCAAATGTCATGATATGAGAAATGATTTATTTGCT 5520
 Qy 5521 taagttatcaaaaatagatggtgatatatacatcttgatcaaaagaagaagaattgaat 5580
 Db 5521 TAAAGTTATCAAAAATAGATGGTGATATACATTTGATGATGATGATGATGATGATGAT 5580

Db 5521 TAGTATATCAAAAATVAGCATGATATCATATGATGATGATGATGATGATGAT 5580
 Qy 5581 accaccaaaatcatttctatcogaatgttcaagtgatgcgtcatttgcacagaattga 5640
 Db 5581 ACCACCAAAATCATTTGTTATCGAATGTTTCAAGATCGCATTTGTCACACAGATTGGA 5640
 Qy 5641 tagaaatcctattagtggaacttgatttatgtgaatggaatgttggtgaatggcgac 5700
 Db 5641 TAGAAATCTTATTAATGGAATCTTTGATTTATGATGATGATGATGATGATGATGAT 5700
 Qy 5701 caaaaaacaaagcgtcatagcaaaagctcagcagcttgtaaatgttgctcaatla 5760
 Db 5701 CAAAAAACAAGCGTCATAGCACAAGCTCAGACCTTCTCAAAATGTTGGCTCAAAATTA 5760
 Qy 5761 tacaatgttgaaagctatcgaataaataaaccattlaatggaattggttgaagtag 5820
 Db 5761 TACAATGTTGAAAGCTATACGAATTAATAAACATTTAATGATTTGGGTTTGAAGTAG 5820
 Qy 5821 taagaatactgtcatagcaaaacagctgtaattgaatgttggaataaactgtt 5880
 Db 5821 TAAGATCATTTGTCATCAAGACACCAAGCTGTGATTTAAAGTTTGAGAAATTAATG 5880
 Qy 5881 taccacacatgcagccaatagatlatcgtataagtttctacgcaatgataatga 5940
 Db 5881 TCACCCACATCGACCAATAGATATCTGCTATTAAGTTTCTACGCCAATGATCAATGATA 5940
 Qy 5941 agtatttccaatactctatggaagaacaaatgataatgaagcgaattgtaagtaagt 6000
 Db 5941 AGTATTTTCAATATCTTATGTAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAGTG 6000
 Qy 6001 tctaagctgcttaaatcaaaagcctgcttgaagtgatgaataaagcgttgaagcaga 6060
 Db 6001 TCTAAGTGTGCTGTAATTCACAAAGCATTTGCTGAGAGGTATGATTAACCGTTAGACTGA 6060
 Qy 6061 agatbaatacaaacctgatacaaaatgcaataaagcgaataaagtggaattatcaat 6120
 Db 6061 AGATBAATCAAACTGATACCAAAATGCAATTAACGCGAATAAGTATGATTTATCTACT 6120
 Qy 6121 attacgttaatgctcaaatcaagggaggttggttggtgactatgttggtgcagaact 6180
 Db 6121 ATTATCGTAATGCTCAATCAAGCGAGGTGTTGTTGCTGACTATTTGTTGTCACAAACTG 6180
 Qy 6181 atcaatgaataatgatttattatgagaatggaatatttccatcaacacatccagtgat 6240
 Db 6181 ATCAATGAATAATGATGTTTATTAATGAGAAATGCAAAATTTTCCATCAACATCGATGAT 6240
 Qy 6241 gacgaactaaactatattgtgtgataaaataagggatagaataccacatccagaa 6300
 Db 6241 GACGAACCTAAACTATATTTGTTAGTAAATAAGGGATGAAATACCAATCCAGAA 6300
 Qy 6301 tacaagagatagaagggaggtttcaatataatcttgtaataaactcgttcc 6360
 Db 6301 TATCAACAGATAGAAAGGAGAGTTTCAATATATATCTTGTGAAATTAATATCTGTTCT 6360
 Qy 6361 aattcactatacaacactagaacgtgtacaacgctcaatctcaggaagaagaattatatt 6420
 Db 6361 AATTCACTATACACACTAGACAGCTGTACAGCTCAATCTCAAGTAAGAAAGTTATATT 6420
 Qy 6421 ccaatca 6426
 Db 6421 CCATCA 6426

RESULT 2
 AF050215 6980 bp DNA 01-JUL-1998
 LOCUS
 DEFINITION Candida albicans Tca2 retrotransposon gag polyprotein (gag) and pol
 ACCESSION AF050215
 VERSION AF050215.1 GI:3273716
 KEYWORDS
 SOURCE Candida albicans.
 ORGANISM Candida albicans

Db 747 CGTCGATATCAAGAAAGCGTGTGAATAATGAAATGAATTTAAATTTCCACGCTGG 806
QY 601 ttcccccgaatttttggaaatttttgccttgactataacttaagaataaacttaagttta 660
Db 807 TTACCCCGAAGTTTGGAAATTTTGTGCTGACTTAAATCCCTAAAGAAATAATTCAGAGTTAA 866
QY 661 aaaggtlaaggaattatttacttggttggttttcaaaatggttttcaagtccttt 720
Db 867 AAAGGTAAAGGATTTATTTACTGGTGTGTTTACAAATGTTGTTTACAGTCCATTTT 926
QY 721 tgaatgttcagattgactaagtttcttaagctcaaaagcacttgcacaagaagcaaa 780
Db 927 TGATAGGTTCAAGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 986
QY 781 cttaataaagcgtcgtatgaatgcttacttaataacttaagaattatacattactgtaa 840
Db 987 CTTAATCAAGGCTGCTTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1046
QY 841 gactctgctgaagtttctgaacgttgaacataagttagttggtttgctataacttcata 900
Db 1047 GATCTGCTGAAGTTTGTAAAGCTGTGAACATGAGTTAGTGTGTTGCTATTAACCTTCCATA 1106
QY 901 ttctgctcaggttgaagaagaacttgaagaataacttcaacaacttcaaaagttgctga 960
Db 1107 TTTCGTCAGAGTGAAGAGAACTTGAAGAAATCTACAAACACTTCAAAACGTTGTGA 1166
QY 961 tgaatgttcaggttgaagaagaacttgaagaataacttcaacaacttcaaaagttgctga 1020
Db 1167 TGAGTATGTCGAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1226
QY 1021 gaaatcaagagcttgaagtttcttgaataatcaatgcttcaacttcaacttcaagttg 1080
Db 1227 GAATCAGAGGCTTAAAGTTGTTTGAATATTCATGCTCATCTACTACTCAAAAGTGAT 1286
QY 1081 tcaagcttgaagaagaacttgaagaagaacttgaagaagaacttgaagaagaacttgaaga 1140
Db 1287 TCAGAGTGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1346
QY 1141 tgaatcaacttgaatgcttgaatgcttgaatgcttgaatgcttgaatgcttgaatgct 1200
Db 1347 TGATCATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1406
QY 1201 tatgttgaatgcttgaatgcttgaatgcttgaatgcttgaatgcttgaatgcttgaat 1260
Db 1407 TATGTTAGTTTACAGAGTGAAGAGAACTTGAAGAGAACTTGAAGAGAACTTGAAGAG 1466
QY 1261 acaatctcaaaagaagaacttgaagaagaacttgaagaagaacttgaagaagaacttgaaga 1320
Db 1467 ACATCTCAAAAGAGAACTTGAAGAGAACTTGAAGAGAACTTGAAGAGAACTTGAAGAG 1526
QY 1321 taaagaatgaagttgaagaagaagaagaagaagaagaagaagaagaagaagaagaaga 1380
Db 1527 TAACGATTAAGAGTGAAGAGAACTTGAAGAGAACTTGAAGAGAACTTGAAGAGAACT 1586
QY 1381 tgcgttcttataatgcttgaatgcttgaatgcttgaatgcttgaatgcttgaatgcttga 1440
Db 1587 TGCCTGCTTCAATTAATGCTTGAATGATGATGATGATGATGATGATGATGATGATGATGAT 1646
QY 1441 agaaatctcaatcttcaatcttcaatcttcaatcttcaatcttcaatcttcaatcttca 1500
Db 1647 AGAAATTTTCATCTTCAATGCTTGAAGAGAACTTGAAGAGAACTTGAAGAGAACTTGA 1706
QY 1501 taaagaatctgaatgcttgaatgcttgaatgcttgaatgcttgaatgcttgaatgcttga 1560
Db 1707 TAACAGATTAAGAGTGAAGAGAACTTGAAGAGAACTTGAAGAGAACTTGAAGAGAACT 1766
QY 1561 attgcgaatgcttgaagaagaagaagaagaagaagaagaagaagaagaagaagaagaaga 1620
Db 1767 ATTGCTGAATGTTAAGAGAGCAATTAAGAGTTGCTGCTGAGAGGCTGACATTTGA 1826
QY 1621 agcagaatgcttgaatgcttgaatgcttgaatgcttgaatgcttgaatgcttgaatgct 1680
Db 1827 AGCAGATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1886

QY 1681 tacatgtatttaccagaagaagttccttcaatcttgaatgcttgaatgcttgaatgcttga 1740
Db 1887 TACATGTTATTTACCAAGAAAGTTCTTAACTTGTGATTTGAAGCAAAATGGAAGAAC 1946
QY 1741 agaatlaaagtttcttacttcaagaagaagaagaagaagaagaagaagaagaagaagaaga 1800
Db 1947 AGGATTTAATGTTCTTATTAAGAAATCAAGTGTGATTTAAACCAAAATGAGCTCC 2006
QY 1801 tactatttgccttcaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaaga 1860
Db 2007 TACTATTAATGCTTCAAGAGAAATGCTGATGATTTATTTAGGCTCTCATTCAGTGA 2066
QY 1861 agaatcttgaagaatgcttgaatgcttgaatgcttgaatgcttgaatgcttgaatgct 1920
Db 2067 AGAATCTTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2126
QY 1921 ccaagatgaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaaga 1980
Db 2127 CCAAGATGACAAAGATTAATCAAGTATGATGATGATGATGATGATGATGATGATGATGAT 2186
QY 1981 tagttcgaagcatlaaataatcttgaagaagaagaagaagaagaagaagaagaagaagaaga 2040
Db 2187 TAGTCTGAGCATTAATTAATTTCTTGACGAGGTTGATTTTAAATGTTGAATTTTC 2246
QY 2041 cccatagagattgagaagaatgcttcaagaagaagaagaagaagaagaagaagaagaagaaga 2100
Db 2247 CCAATATGAGAGTTGAACAAATTTGATCAACAGTGAAGATGAAGATTAATTAATTTTCA 2306
QY 2101 ttgtatgcaaatcaatgaatgcttgaagaagaagaagaagaagaagaagaagaagaagaaga 2160
Db 2307 TTGATGTCAAATCATATGTCATGAGAAATCTGTTGTTTACAAAATCCAGGCTCT 2366
QY 2161 cgtactcaacacttcaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaaga 2220
Db 2367 CGTACTGACACTTCAAGAGAGAGCTTCAAGATGCTGATGTTGAAGGATGCTATTT 2426
QY 2221 atcgaatgcacaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaaga 2280
Db 2427 ATCGAATGCCAAAGAGAGAGAGTCAATCATCATTCACAAAAGAGCTTGAAGAGAGCA 2486
QY 2281 tgaagaactcaatgcttgaatgcttgaatgcttgaatgcttgaatgcttgaatgcttga 2340
Db 2487 TGAGAGACTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2546
QY 2341 aacgtcttcttgaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaaga 2400
Db 2547 AACGCTGTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2606
QY 2401 ggtlaaagatcttcaatcaagaagaagaagaagaagaagaagaagaagaagaagaagaagaaga 2460
Db 2607 GGTAAAGATGCTTAATTAACAGATTAAGATGATGATGATGATGATGATGATGATGATGAT 2666
QY 2461 ggcatacttcaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaaga 2520
Db 2667 GGCATATCTTCAAGAGTGAATGCTTCAAGATGATGATGATGATGATGATGATGATGATGAT 2726
QY 2521 cgtatattgagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaaga 2580
Db 2727 CGGATTTTGGAGGAGACTATAGCGCATATCTGCTGAGGTTATAGGCTTGGCGAGGT 2786
QY 2581 tgttaataatgcttcaatcaagaagaagaagaagaagaagaagaagaagaagaagaagaagaaga 2640
Db 2787 TGTATTAATTAATTTTACAAAGATTTACAGATGCTTGTGACACTTGTGTCACAAAT 2846
QY 2641 actcaagttgcttcaatgcttgaatgcttgaatgcttgaatgcttgaatgcttgaatgct 2700
Db 2847 ACTCAAGTTGATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2906
QY 2701 tgcgttcaatcaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaaga 2760
Db 2907 TCGTTCACTCAAGGAGCAAAACCCCTTATGTTGCTATTTATCAATTAAGTGAAGGAAATTT 2966

QY	2761	ctaacggttccctcttggatcacgtctgtgctgtacattgaatgaatgcctgaagaaga	2820
Db	2967	CTACCGGGTTCCTTTGGCATTGCATGATGTGTGCTTACATTGATATGCCATCCAAAAGAA	3026
QY	2821	ccgttacagagatcatcaatcaataaagagagcctcttcacatcatggtgtcgttgatgg	2880
Db	3027	CCGTTACGAGATTACATCAACTTAAGAAGAGCCTTTCATCATCATGAGGTGCTGTGATTTGG	3086
QY	2881	ctaagctagagattggttttagttatctacgtgttgcctaaataatgtggtgtatattat	2940
Db	3087	CTACGCTAGCCATTGTTTTAACTTATATACGTGTTCATAAAAAATATGCGGTATATATAT	3146
QY	2941	ccttaagccctaatgtccgcatatctgcgaagcagctacgaagttatcaaccctcatccaana	3000
Db	3147	CCTTAGCCCTATGTCCGCTATATTGCCAGAGCTATGAGTATTATACCTCATCTCAAAA	3206
QY	3001	cctatccatcacacctatgttcaacaactgttccctcctcgtgaagttatccaggaagaca	3060
Db	3207	CTTATCCCACTACACTATAGTCACACATTTGCTCATGTGCGAAGGATATCCAGGGAAGGCA	3266
QY	3061	actgagcgctcagtaagcaggtgacgcggaacaatgtgtgaagtgaaatgatgaacatacaa	3120
Db	3267	ACTGGGCGCTCATGATACGAGAGTACGCGGAGACATATGTGGAAAAGTGAATATACATACAA	3326
QY	3121	tgaagtgatgcacatgcccagaagagtcataatcagttcaagccagcatcgtttacttaac	3180
Db	3327	TGACGTGATGCACATGCGCCAAAGAGATCATATTTCAGTTACACCGACATCGTTACTTTAAC	3386
QY	3181	tacgggttaaggtttcttaacgaatatgttataatgatgatccagtaagttacccattga	3240
Db	3387	TACGGGTACAGTTCTTAACGAATATGTTATTAATGATGATCCCATACGAATACCATTTGA	3446
QY	3241	gaatcccgatgatttttctaacacctcttcaactacactagaagaatcaacagataltggtalc	3300
Db	3447	GAATCCCATGATTTTCTTACACCTCTTCAACTACAGAAATACACAGATATGTGATTC	3506
QY	3301	cgaagttaaatcagatagagaatccttaaacccagctctccagagcttaacacttggagata	3360
Db	3507	CGAATTAATAATCGGATGAGAACTCTAAACCCAGTCTCCAGAGCTTAACACTGGGGATAA	3566
QY	3361	lccggtgtctaaacctccctcaacttggtaaccgagacttcagtaatgaaggagcttaaga	3420
Db	3567	TCCGGTGTCTAAACCTCCTCAACTTGGTACCGAGACTTCAGTAAATAGGAAGTCTAAAGA	3626
QY	3421	gctctattcaacaacacaaagaagccctctccaatccagggagggagccataaagcct	3480
Db	3627	GCTATTATCAACAACACAAAGAGCGCCCTTCCATCCAGGGGAGGACCATATAACGCCCT	3686
QY	3481	ggaatctactgtcaaggttgcacatcaacaacacccagactgtaactccgcttcogga	3540
Db	3687	GGAATCTACTGCTCAGTGTGAGTACATCAACACACCCAGACTGTGATCTCCGCTTCGGA	3746
QY	3541	ggagtcacaaattgtcgaagacagatcatcttcggtgtcgaagttgttaagaacagctctc	3600
Db	3747	GGAGTCAAAATTTGTCCAGGACAGATCATTTTCGTGTGAGCTGTGTTAAAGAAACAGTCTC	3806
QY	3601	agaagaattggaactcttcgtactccacagaanaactgaagtcggaagaatgaagaacgaatcc	3660
Db	3807	AGAAGATTGGCATCTTCTGCTACCCAGAAACTAGTGTGCTGAAGATGAACAGAAATTC	3866
QY	3661	ctcgttacttggctaaataagaatcgggttaactgaagaaataatagatgaaggaggaataatlc	3720
Db	3867	CTCGTTACTGGCTAATAGATCGGGTAACTGAAAAAAATGATGAGGAGAAATATTTTC	3926
QY	3721	atttcgggggggtgatgatgatctctgtcgtgatcaactcaaatggttgagcaatcctaatgt	3780
Db	3927	ATTTCGGGGGGGTATGATGTGATCTGTGCTGATCAACTCAAAATGTTGACGAATCTAATGT	3986
QY	3781	tgaacaacagagatctgtttaaagtcacatccaagtcgaagattctcaagaagggaagaat	3840
Db	3987	TGAAACAGAGAGATGCTGGTAACTCCATATTCAAGACGAAGTTTCTCAAGAGGGAACAT	4046
QY	3841	acttaatgaacaaactgatalagttgtgatactgtgtgctaagaattatgagaatgaanaaat	3900
Db	4047	ACTTATATCAAACTGATATAGTTGATCTGTTGCTTAAAGTTATTTAGAAATGAAAAAT	4106
QY	3901	ctctcccttaattcatatagatgatcatcactgaactgtctacagactcgggaatgatag	3960
Db	4107	CTCTCCCTATTATTCATTAGATGATGATCAATACGAACTTGTACAGACTCGGGAATATATAG	4166
QY	3961	caattcaacgaatccgacatccaatccaataagaanaatgaanaatccacagttgattatgaag	4020
Db	4167	CAATTCAACGAATACTCCGATTCATCAATGAAATAATGAAATATCCACAGTATTATAGAA	4226
QY	4021	aaatctgaataatcccaaaaaacatctgaagaatccttgcctgcataagaagatttgatga	4080
Db	4227	AAATCTGAATATATCCAAAACATTTGAAGATATCCTTGCTGATTAAGATTTGGATGA	4286
QY	4081	atttgaacgtaataatgttggatgaattggaatgtgataatgaagatgaacttgcga	4140
Db	4287	ATTTGAACGTTATATGTGTGATGAATTTGAATAATGATTAATGACATATTCCTGTA	4346
QY	4141	agctaatccactacagaatgaanaataatgatttcaagatgaatgaagatttataataa	4200
Db	4347	AGCTATCTCACTACAGATGAATAATTAATGATGTTCAATGAATGATGAGTGTGATATA	4406
QY	4201	tcaatgactgtcaagcagaagaagaataacacatcttggagaagaagttaaagaaaaaat	4260
Db	4407	TCAATGACATGTCCAGAGAGAAAGAAATACCATTTTGAATAAATAAGTTTAACAAAAAT	4466
QY	4261	tgtgtgtactaaacatcaacttgaacttgtaacactgcgaagaagaacaaatcgaagttgtaa	4320
Db	4467	TGCTGTACTTAACATTTCACTTGATATACACATGATCCAGAACGATTAACAGTCTTAA	4526
QY	4321	tacttgtgaacccaagagaatcgaacccaagaagaagaaggtgcctacatcgttgaat	4380
Db	4527	TACTGTGTAACCAAGAAATATCGAACCCAGAAAGAGAGTGCCTATCTGATGAAAT	4586
QY	4381	aaacaagaatccgaataatgaagttacacatagttaagtaagaagaatgttgtaacggtttaa	4440
Db	4587	AAACAAAGATCCGATATCAAGTACCATATGTTTACAAGAAAGTGTGAACGGTTATAA	4646
QY	4441	cccccaaggtatttaccatgagtggttcgttcaacaaatccgactaataatgataccggatgat	4500
Db	4647	CCCCAAGAGGATTATTCATGCGCGTCGTCACAAATGACATATATATCCGGGATGGAT	4706
QY	4501	aaagtcgaatgactgtgaactcagaagaatttagatcaaaagaatgtttcagaagaattcc	4560
Db	4707	AAAGTCAATGATGTCTGACTAGAGAAATTTAGTCAAAAGATGTTTACGAAGAAGTTC	4766
QY	4561	aattcccccgglttgaagccctacatcatatggttggatgatcatacttgaagaaattgattc	4620
Db	4767	AATTCACCGGTTGGAAGCCTATATCTATGATGTTGGGTACATATTCGAAAAATTTGATTC	4826
QY	4621	tctcaaaagtgttgttgcgaaatcaacglttgttgcctcagtcagcaacgaacaaagaagaa	4680
Db	4827	TCTCAAAAGGTTGTTCGGAATACAGTGTGTTGTCATATGCGAACGACAAAAAGGAAA	4886
QY	4681	attgatcatgaacctttatgttgaatgttgaactccgttatagatcttcttgactaagat	

[illegible]

QY	6061	agataatcaaaacctggtacaaaatgcaataagcgagaataatggtgattatcaattact	6120
Db	6267	agatatactaaacactgtgtatcaaaaatcattaacgcgagatpactggtttatcattact	6326
QY	6121	atatacgtatgacccaatcagggagtgcttggttttgacattttgtccagaactg	6180
Db	6327	attatccgaatgccatcaccgggagtgtttggtttgtccactatTTTTGTGTGAGAAATG	6386
QY	6181	atcaatggaanaatgctgttattatgagaaatggaanaattttccatcacacatcagtgat	6240
Db	6387	atccatggaanaatgattggttatpattatgagaaatggaaaattttccatcacacatcagtgat	6446
QY	6241	gacggaactaaactatattgtctgatatataataaaggatgtaaatcaaccacatcccgaa	6300
Db	6447	gacggaactaaactatattgtctgatatataataaaggatgtaaatcaaccacatcccgaa	6506
QY	6301	tatacaagaataagaaggaggagttcaatatatacttggaaataataacttcgcttc	6360
Db	6507	tatatcacagatgataagacggagagtttcaattatpattcttggaaataataacttcgcttc	6566
QY	6361	aattccatatacaacaactaagctgtagcagcgtccaatccaggtlaagaaagttatatt	6420
Db	6567	aattccatatacaacaactaagcgtgtagcagcgtccaatccaggtlaagaaagttatatt	6626
QY	6421	ccatca 6426	
Db	6627	ccatca 6632	

RESULT	3			
LOCUS	CAL251464/c			
DEFINITION	CAL251464	4872 bp	DNA	PLN 10-DEC-1999
ACCESSION	U00000			
VERSION	AJ251464			
KEYWORDS	endo-1,3-beta-glucanase; eng1 gene; retrotransposon.			
SOURCE	endo-1,3-beta-glucanase; eng1 gene; retrotransposon.			
ORGANISM	Candida albicans			

REFERENCE
1 (bases 1 to 4872)
AUTHORS
Esteban, P. F., García, R., Ríos, I., Vázquez De Aldana, C. R. and del Rey, F.

TITLE	Cloning and characterization of the endo-1,3-beta-glucanase-encoding gene CAEN1 in the yeast <i>Candida albicans</i>
JOURNAL	unpublished
REFERENCE	2 (bases 1 to 4872)
AUTHORS	del Rey, F.
TITLE	Direct Submission
JOURNAL	Submitted (07-DDC-1999)
FEATURES	Genetica, Universidad de Salamanca, Campus Miguel de Unamuno, Salamanca, 37007, SPAIN Location/Qualifiers

```

source      /organism="Candida albicans"
            /db.xref="taxon:5476"
            1..4440
            /organism="Candida albicans"
            /strain="ATCC64385"
            /db.xref="taxon:5476"
            711..4148
            /gene="eng1"
            711..4148

```

```

/codon_start=1
/transl_table=12
/evidence=experimental
/product="endo-1,3-beta-glucanase"
/protein_id="CAB62579.1"
/db_xref="GI:6562334"

```

LOCUS	AF030556	408 bp	DNA	PLN
DEFINITION	Candida albicans retrotransposon pcal ITR, and GAG gene, partial cds.			
ACCESSION	AF030556			
VERSION	AF030556.1	GI:3273502		
KEYWORDS				
SOURCE				
ORGANISM	Candida albicans. Candida albicans			
REFERENCE	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; mitosporic Saccharomycetales; Candida.			
AUTHORS	1 (bases 1 to 408) Goodwin,T.J.D. and Poulter,R.T.M.			
TITLE	Temperature- and strain-dependent expression of the pcal retrotransposon of Candida albicans			
JOURNAL	unpublished			
REFERENCE	2 (bases 1 to 408) Goodwin,T.J.D. and Poulter,R.T.M.			
AUTHORS	Direct Submission			
TITLE	Submitted (21-OCT-1997) Department of Biochemistry, University of			
JOURNAL	Otago, P.O. Box 56, Dunedin, New Zealand			
FEATURES	Location/Qualifiers			
SOURCE	1..408			

BASE COUNT	143 a	63 c	75 g	127 t
ORIGIN				

Query Match	6.28:	Score 400:	DB 12:	Length 408:
Best Local Similarity	98.88:	Pred. No. 8-be-55:		
Matches 403: Conservative	0:	Mismatches 5:	Indels 0:	Gaps
QY 1	ttgttggttggaacattttgtgtcgaagaactgatacatgaataatgatgttattga	60		
Db 1	TTGTGGTTTGTGCACATATTTTGTGTGAGAAACTGATCAITGAAATATATGTATTATGA	60		
QY 61	gaatgaaaaattttccatccacacatcaggtgagtgacagaactaaactatattgtagt	120		
Db 61	GAATGAAAAATTTTCCATCCACATCAGGTGATGACAGAACTAAACTATATTTGTAGT	120		
QY 121	ataaataagggatgagaattaccaacatcccgagaatatacagaatataagggagaggtc	180		
Db 121	ATTAATTAAGGGATGAAGAAATACCACATCCCGAATATATCAGAAATGAAAGGGAGGATT	180		
QY 181	tcaatatatattctgtgataataacttcgttcttaattcacctatacacaactagaagtc	240		
Db 181	TCAAATATATATCTGTGATATATATACATTCGTTCTTAATTCACATATACACAACTAGACGTGT	240		
QY 241	acaagcctaacttcgaaggaagaagttatataccatccagattagagaagtcgatatgcat	300		
Db 241	ACACGCTCAATCTCGATGAAGAAGTTTATATTCACATCAATATAGAAAGTCAAAAGGAT	300		
QY 301	aatcatcttcgtcccaaatltagcgttgtataaattcagtcctcagatttgtattatgcat	360		
Db 301	AACCAATTCGCCCCCAATATAGCGTGTATTAATTCAGTCCCAATTTGTTATTTGATT	360		
QY 361	gatagtttcgaagtttagaagtgacagaatttcacaagaatgattccgc	408		

QY 5659 aactltgattatgtatggaatttctgtgcaatgtggcgaacaaacaaacggatcat 5718
 Db 1321 GACAAAGCATTTTGTAAATGAGATCTAGTTGCATGGGCAACAAAGAAACAAATAAT 1380
 QY 5719 agcacaagctcagcagctgtgaaatgtgtcttaatatatacaatgttgaaatcat 5778
 Db 1381 CACCAATAGTTCACCTCTTGGCAATGCTTGATTAATATATACATATTTGAAGCAT 1440
 QY 5779 cgaataaaaa 5790
 Db 1441 TGATTTCAGAAA 1452
 RESULT 6
 CDS07BD 1017 bp DNA STS 12-JAN-2001
 LOCUS 17 end of clone BC0AA012E08 of library BC0AA from strain CBS 767 of
 DEFINITION Debaryomyces hansenii, sequence tagged site.
 ACCESSION AL437951 GI:12221364
 VERSION 1
 KEYWORDS
 ORGANISM Debaryomyces hansenii.
 Debaryomyces hansenii.
 Debaryomyces hansenii.
 Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 Saccharomycetales; Saccharomycetaceae; Debaryomyces.
 REFERENCE 1 (bases 1 to 1017)
 AUTHORS Lepingle, A., Casaregola, S., Neuveglise, C., Bon, E., Nguyen, H.,
 Artiguenave, F., Wincker, P., and Gaillardin, C.
 TITLE Genomic Exploration of the Hemiascomycetous Yeasts: 14.
 JOURNAL Debaryomyces hansenii var. hansenii
 PUBMED FRBS Lett. 487 (1), 82-86 (2000)
 11152889
 2 (bases 1 to 1017)
 REFERENCE Souciet, J.L., Aigle, M., Artiguenave, F., Blandin, G.,
 Bolet, J., Fukuoka, M., Bon, E., Brothier, P., Casaregola, S.,
 de Montigny, J., Dujon, B., Durand, P., Lepingle, A., Llorente, B.,
 Malpertuy, A., Neuveglise, C., Ozier, K., Papadimitrakis, F., Potier, S.,
 Saurin, W., Tekala, F., Toffano-Nioche, C., Wesolowski-Louvel, M.,
 Wincker, P., and Weissenbach, J.
 TITLE Genomic Exploration of the Hemiascomycetous Yeasts: 1. A set of
 yeast species for molecular evolution studies (1)
 JOURNAL FRBS Lett. 487 (1), 3-12 (2000)
 11152876
 3 (bases 1 to 1017)
 REFERENCE Genoscope.
 AUTHORS Direct Submission
 JOURNAL Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage,
 2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex - FRANCE (E-mail :
 seqref.genoscope.cns.fr - Web : www.genoscope.cns.fr)
 COMMENT This STS is part of a random genomic sequencing program of thirteen
 yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
 exiguus, Saccharomyces servazii, Zygosaccharomyces rouxii,
 Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
 lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
 angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,
 Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
 5 kb were prepared and both extremities were sequenced. See
 keywords for description of this sequence and for the sequence of
 the other extremity of this insert.
 FEATURES
 SOURCE
 1. 1017
 /organism="Debaryomyces hansenii"
 /strain="CBS 767"
 /variety="hansenii"
 /db_xref="taxon:4959"
 /clone="BC0AA012E08"
 /clone_1ib="BC0AA"
 /note="end : 77"
 <1..>1017
 misc.feature
 /note="part of putative transposable element"
 /evidence=not-experimental
 BASE COUNT 367 a 177 c 158 g 313 t 2 others

ORIGIN
 Query Match 2.7%; Score 174.6; DB 53; Length 1017;
 Best Local Similarity 53.2%; Pred. No. 1.1e-18;
 Matches 412; Conservative 1; Mismatches 360; Indels 2; Gaps 2;
 QY 4444 caaaggattatgaatgcgcgtgtcaacaaatcgaactataatgacccggatgataa 4503
 Db 244 CACTCGGAAATATTAATGATGATATGAAAGTCGATTTAACAATTAACAATCAATCAAT 303
 QY 4504 gtcaatgaatgtctgaacaggaatattagatcaaaagtgttcaagaagaattccat 4563
 Db 304 GTCAATACACCGGATTTGATGATTTTAAAGAAATATTAATATTAATATTAATATTAAT 363
 QY 4564 tccacacggtgtgaagcctatactatgattgtgtgtgtgtgtgtgtgtgtgtgtgtgt 4623
 Db 364 TCCCTAAATATGTCAAACCATTCACATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 423
 QY 4624 caaaggt 4683
 Db 424 CAAGAAGTTCAGTATTAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 483
 QY 4684 gattatgaacccctttagt 4743
 Db 484 ACACATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 543
 QY 4744 gcaaatataggt 4803
 Db 544 TACAGGATACAGT-TGATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 602
 QY 4804 tctaatgtctctatcaccatcaacatcaatctgtcttcttcttcttcttcttcttcttct 4863
 Db 603 TTTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 662
 QY 4864 ttgaagaataaacacatgt 4923
 Db 663 TTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 722
 QY 4924 ttgaaggt 4983
 Db 723 GTATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 782
 QY 4984 acaaatgt 5042
 Db 783 TGCCATTTAGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 842
 QY 5043 tatatgt 5102
 Db 843 TATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 902
 QY 5103 atcaatgt 5162
 Db 903 TTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 962
 QY 5163 ttgaatgt 5217
 Db 963 TTGAATTTCAATCAATTTCAATTTCAATTTCAATTTCAATTTCAATTTCAATTTCAAT 1017
 RESULT 7
 AX073190 1888 bp DNA PAT 25-JAN-2001
 LOCUS Sequence 301 from Patent WO0102550.
 DEFINITION AX073190
 ACCESSION AX073190
 VERSION AX073190.1 GI:12583390
 KEYWORDS
 SOURCE
 ORGANISM
 Candida albicans.
 Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 Saccharomycetales; mitosporic Saccharomycetales; Candida.
 REFERENCE 1 (bases 1 to 1888)
 CONTRERAS, R.H., de Backer, M.D., Luyten, W.H., Malcorps, I.K.,

BAC clone F114 is from Arabidopsis thaliana chromosome 1. The orientation of the sequence is from SP6 to T7 end of the BAC clone. Genes were identified by a combination of several methods: Gene prediction programs including GenScan+ (Chris Burge, <http://CCR-081.mlt.edu/GENSCAN.html>), GeneMarkHMM (Mark Borodovsky, <http://genemark.biology.gatech.edu/GeneMark/>), Glimmer4 (a variant of Glimmer, see Mihela Pertea, <http://www.tigr.org/softlab/glimmer.htm>), and GeneSplicer (Mihela Pertea and Steven Salzberg, contact mpertea@tigr.org), searches of the complete sequence against a peptide database and the plant EST database at TIGR (<http://www.tigr.org/tdb/tgi.shtml>). Annotated genes are named to indicate the level of evidence for their annotation. Genes with similarity to other proteins are named after the database hits. Genes without significant peptide similarity but with EST similarity are named as unknown proteins. Genes without protein or EST similarity, that are predicted by more than two gene prediction programs over most of their length are annotated as hypothetical proteins. Genes encoding tRNAs are predicted by tRNAscan-SE (Sean Eddy, <http://genome.wustl.edu/eddy/tRNAscan-SE/>). Simple repeats are identified by RepeatMasker (Arian Smit, <http://ftp.genome.washington.edu/RM/RepeatMasker.html>).

FEATURES

Source

1. 80561

/organism="Arabidopsis thaliana"

/cultivar="Columbia"

/db_xref="taxon:3702"

/chromosome="1"

/clone="F114"

join(762..801,1066..1183,1260..1333,1447..1570,1805..1987,2071..2312)

/gene="F114.1"

/note="similar to 40S ribosomal protein S7 homolog GI:5532505 from (Brassica oleracea)"

join(1109..1183,1260..1333,1447..1570,1805..1987,2071..2190)

/gene="F114.1"

/codon_start=1

/product="40S ribosomal protein S7 homolog, putative"

/protein_id="AAG60128.1"

/db_xref="GI:1259781"

/translation="MESAQHKIKKKEGVELDEOVAQAFDLENTNOELKSELKDLVNSAVQDISGKRAIVNVPRLKRAIKRIHYRLVELEKESGKDVILARRIVRPPKSGSAKRRPKRLTSHVAILDDVLPALVIGKTRIRYLDGFKIMKVELDPKRRNTEKVAFAFAVKKLKGKDVFEFPITEA"

complement(join(<2677..2964,3055..3303,3412..3508,4553..4771,4862..5230,6284..6492,6561..>6815))

/gene="F114.2"

/note="similar to 40S ribosomal protein S7 homolog (Arabidopsis thaliana)"

complement(join(2677..2964,3055..3303,3412..3508,4553..4771,4862..5230,6284..6492,6561..>6815))

/gene="F114.2"

/codon_start=1

/product="terpene cyclase, putative"

/protein_id="AAG60130.1"

/db_xref="GI:12597819"

/translation="MEATRVGKLGSKLSHSQNAFPAPKLSRPFPLTFPGKAHLDPKATTHPLAFDEENRERKNGSPSEGHQFSAHVDISEMDALEREALPKRYADMLISESSKRIIFLYLVSLGLAFHEDEIKESLDEGLOKIEEMASSEDLEKFDNGKEKELADAKGILSLYEAAHMGTTDDYLIDALSETTLVMEASLASGCKINTLSRIRKALDPOKRMKEIIVAKYIOFYEEEDCDKTLKFAKLNKRFLOLHYLDEKTLISKWKKDOFKSKLPPYFEDRLVCHFASTGCPKRYARAFLEKIKITVOITDITDTSYASIGVEESLADTERDPPDHADGIPDLTKSVKVFVTFQEFERKCKRSRLINLQVAKVKAHLPSFDEYLDVAGLELAISFTAGILGMEVCKPEAEVEMLSRDLKLVGVITKRLNLDIFEGEDMRGYTNSINCKKOYQTEEAIRKLQIIVADGKKMNEBFLKPIVPOVPRVIVLIDTLRAANASYEKDDEFTRGHEHLKNCITSIYDL"

complement(4261..4281)

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

/rpt_family="AT_rich"
complement(5501..5529)
/rpt_family="AT_rich"
complement(5593..5617)
/rpt_family="AT_rich"
complement(5735..5755)
/rpt_family="AT_rich"
complement(6117..6171)
/rpt_family="AT_rich"
7685..8673
/gene="F114.19"
/note="pseudogene, putative retrotransposon polyprotein"
/pseudo
complement(join(<9069..9353,9447..9695,9795..9942,10061..10279,10375..10753,11586..11862,11946..>12200))
/gene="F114.3"
complement(9069..12200)
/gene="F114.3"
/note="similar to terpene cyclase GI:9293912 from (Arabidopsis thaliana)"
complement(join(9069..9353,9447..9695,9795..9942,10061..10279,10375..10753,11586..11862,11946..>12200))
/gene="F114.3"
/codon_start=1
/product="terpene cyclase, putative"
/protein_id="AAG60132.1"

/db_xref="GI:12597821"

/translation="MEATRVGKLGSKLSHSQNAFPAPKLSRPFPLTFPGKAHLVLKATTHPLAFDEENRERKNGSPSEGHQFSAHVDISEMDALEREALPKRYADMLISESSKRIIFLYLVSLGLAFHEDEIKESLDEGLOKIEEMASSEDLEKFDNGKEKELADAKGILSLYEAAHMGTTDDYLIDALSETTLVMEASLASGCKINTLSRIRKALDPOKRMKEIIVAKYIOFYEEEDCDKTLKFAKLNKRFLOLHYLDEKTLISKWKKDOFKSKLPPYFEDRLVCHFASTGCPKRYARAFLEKIKITVOITDITDTSYASIGVEESLADTERDPPDHADGIPDLTKSVKVFVTFQEFERKCKRSRLINLQVAKVKAHLPSFDEYLDVAGLELAISFTAGILGMEVCKPEAEVEMLSRDLKLVGVITKRLNLDIFEGEDMRGYTNSINCKKOYQTEEAIRKLQIIVADGKKMNEBFLKPIVPOVPRVIVLIDTLRAANASYEKDDEFTRGHEHLKNCITSIYDL"

complement(11212..11236)

/rpt_family="AT_rich"

complement(12289..12333)

/rpt_family="AT_rich"

complement(12467..12519)

/rpt_family="TAAAn"

complement(12725..12779)

/rpt_family="AT_rich"

complement(13575..13612)

/rpt_family="AT_rich"

join(414617..14730,15176..15240,15336..15411,15494..15670,15850..15944,16074..16176,16261..16307,16462..16789,17116..17205,17304..17366,17451..17519,17652..17714,17804..17856,17945..>18119)

/gene="F114.4"

14617..18119

/gene="F114.4"

/note="predicted by genemark hm"

join(414617..14730,15176..15240,15336..15411,15494..15670,15850..15944,16074..16176,16261..16307,16462..16789,17116..17205,17304..17366,17451..17519,17652..17714,17804..17856,17945..>18119)

/gene="F114.4"

/codon_start=1

/product="hypothetical protein"

/protein_id="AAG60133.1"

/db_xref="GI:12597822"

/translation="MSSSETTIDIASRIGVGNRISLKFRIADNLKANIFRAEKVLDIYVWLFFSLAETTPSHRDYTSLSKSNKYLRMLDLVTELEKLPVQORDTELPIKATPKRYNVQAHFANSGLSSAVKSFNSYDHARVNPNGINSQMSRGQFVLNAPLEERKRKMSVNPREFLSKHSILGPGLSAQMPKDYKVOYPSNIDFSFVLPSPQOILVDEKMTNSNDPEKPIVEPSVANSKLOKNTYETLSMTSPEEPEVSNMDETMRALASNTKKNLCTGILAGSLCNKRYITALLIIPQESISDQATNDEEIPRVQKQSLFPLGILHTHPQSCFMSIDVHTHTSYQIMLFEAVALVWAPDQSSRNH

Oy 5033 tattagattatattgtagatattcttattggttggaagttcaaaaaagtattgat 5092
 Db 15881 ATGCGATCTTATATGATGACTTGTATTCACGGGTAACAAATCCAGCATGTTGAA 15822
 Oy 5093 aatttcggatcaattagagatcatttgaattgaagtggttggaataatcaat 5152
 Db 15821 GAATTCAGAAAGAGATACGACAGAGTTCAGATGACGACATTCGATTCCTTAC 15762
 Oy 5153 tattcttgattgaatttcgtlaaaacccgaatcgtttattattctcaagaanaatt 5212
 Db 15761 TATTCGGAATTCAGATTAACAGCAAGCAATGCAATATTCATACACAGAGCTAT 15702
 Oy 5213 ctcaagaataactaagattcaactaagatgactat 5254
 Db 15701 GCTAAGAGGACTTACAGAGTTCAGATGATGACTCAAT 15660

RESULT 13
 LOCUS AC079604/c
 DEFINITION Arabidopsis thaliana chromosome 1 BAC T15M6 genomic sequence,
 complete sequence.
 VERSION AC079604.5 GI:12321249
 KEYWORDS HTG.
 SOURCE thale cress.
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
 Brassicales; Brassicaceae; Arabidopsis.
 1 (bases 1 to 103637)
 Lin.X., Kaul.S., Town.C.D., Bentin.M., Creasy.T.H., Haas.B.J.,
 Wu.D., Maiti.R., Rensing.C.M., Koo.H., Fujii.C.Y., Utterback.T.R.,
 Barnstead.M.E., Bowman.C.L., White.O., Nierman.W.C. and Fraser.C.M.
 Arabidopsis thaliana chromosome 1 BAC T15M6 genomic sequence
 Unpublished
 2 (bases 1 to 103637)
 Town.C.D. and Kaul.S.
 Direct Submission
 Submitted (04-SEP-2000) The Institute for Genomic Research, 9712
 Medical Center Dr, Rockville, MD 20850, USA, cdtown@tigr.org
 3 (bases 1 to 103637)
 Town.C.D. and Kaul.S.
 Direct Submission
 Submitted (19-JAN-2001) The Institute for Genomic Research, 9712
 Medical Center Dr, Rockville, MD 20850, USA, cdtown@tigr.org
 On Jan 19, 2001 this sequence version replaced g1:12280852.
 Address all correspondence to:at@tigr.org

COMMENT
 BAC clone T15M6 is from Arabidopsis thaliana chromosome 1
 The orientation of the sequence is from SP6 to T7 end of the BAC
 clone.
 Genes were identified by a combination of several methods: Gene
 prediction programs including Genscan+ (Chris Burge,
 http://CCR-081.mit.edu/GENSCAN.html), GenemarkHM (Mark Borodovsky,
 http://genemark.biology.gatech.edu/genemark/), Glimmer4 (a variant
 of Glimmer3, see Mihalea Perlea, <http://www.tigr.org/softlab/glimmer4/glimmer4.html>, and
 GeneSplicer (Mihalea Perlea and Steven Salzberg, contact
 mperlea@tigr.org), searches of the complete sequence against a
 peptide database and the plant EST database at TIGR
 (<http://www.tigr.org/tdb/cgi.shtml>). Annotated genes are named to
 indicate the level of evidence for their annotation. Genes with
 similarity to other proteins are named after the database hits.
 Genes without significant peptide similarity but with EST
 similarity are named as unknown proteins. Genes without protein
 or EST similarity, that are predicted by more than two gene
 prediction programs over most of their length are annotated as
 hypothetical proteins. Genes encoding tRNAs are predicted by
 tRNAscan-SE (Sean Eddy, <http://genome.wustl.edu/eddy/tRNAscan-SE/>).
 Simple repeats are identified by RepeatMasker (Arian Smit,
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>).
 Location/Qualifiers

source
 1. 103637
 /organism="Arabidopsis thaliana"
 /cultivar="Columbia"
 /db_xref="taxon:3702"
 /chromosome="1"
 /map="2"
 /clone="T15M6"
 complement(1.. 663)
 /gene="T15M6.26"
 /note="pseudogene, hypothetical protein"
 /pseudo
 complement(1003..1088)
 /rpt_family="AT_rich"
 complement(1091..2595)
 /gene="T15M6.25"
 /note="pseudogene, hypothetical protein"
 /pseudo
 complement(<3359..>4729)
 /gene="T15M6.24"
 complement(3359..4729)
 /gene="T15M6.24"
 complement(3537..4729)
 /gene="T15M6.24"
 /codon_start=1
 /product="unknown protein"
 /protein_id="AAG50699.1"
 /db_xref="GI:12321255"
 /translation="MNTBASKQTTAEGSANDPEPQILSPRSLEKOKKMWISVLC
 IFLVLLGSLVALLNFFVYDQNRDSDDDIQRRTWLQALVONAFPLLPFTIFP
 SFRQNETNRELFSEFRILILYISGLVVAHSKLFAGKLYANGVFTLLSATQIL
 TAFPAIIRNFEFTFMIILISGLIVFSGPERGEPDEEFTYSIOMLTAAVY
 APALSLCEFLCEKVLVTKRKGKRVEMIQICVFPAIVVCVGLFASGK
 FLOGSHRKEGTGYVYVLSIGLAIMOVNAGLGLVYSGVGVGDVHCTSPVLA
 LTVVLAFPMDPEFSMPRTGILAIYVALGSIYTLHRRNKKKWEYLQTEINNDY"
 complement(6149..6169)
 /rpt_family="AT_rich"
 complement(6713..6740)
 /rpt_family="AT_rich"
 complement(7047..7147)
 /rpt_family="AT_rich"
 complement(8659..8680)
 /rpt_family="AT_rich"
 complement(10552..10573)
 /rpt_family="AT_rich"
 complement(10742..10991)
 /gene="T15M6.23"
 /note="pseudogene, putative ribosomal protein"
 /pseudo
 complement(11760..11838)
 /rpt_family="AT_rich"
 complement(12070..17708)
 /gene="T15M6.4"
 /note="pseudogene, putative non-LTR retroelement reverse
 transcriptase"
 /pseudo
 complement(101n<18837..19207,19528..19647,19786..19886,
 20080..20291,20407..20787,20902..20989,21228..21288,
 21891..22125,22758..22835,22952..>23614))
 /gene="T15M6.22"
 complement(18837..23614)
 /gene="T15M6.22"
 /note="contains similarity to cell proliferation related
 protein GI:9931486 from (Mus musculus)"
 complement(101n(18837..19207,19528..19647,19786..19886,
 20080..20291,20407..20787,20902..20989,21228..21288,
 21891..22125,22758..22835,22952..23614))
 /gene="T15M6.22"
 /codon_start=1
 /product="hypothetical protein"
 /protein_id="AAG50696.1"
 /db_xref="GI:12321252"
 /translation="MKRRGHRKGGKSKTINQGNLNESTENAENQTSRHSSEAVEC
 ENSELESKMEVDAPSPFAGSVDLADNIAKSVARVKKTKTSKADPEPNTLLDDIDKS

SSOALEKPEYVPEKEEFPYLERPERPELVNTYRKIKGIRISKSNAVDSSSETEKS
ADTFKVDVVVQKDDTFPEKQSRSSQSLGVGAEIWSVPAELIIRSSSEPOCLVQSGDGA
SLIYIKTILMKHEADPFVWVYVNPALGIDIPYDFDIKIPDMFGITCNNEFKGNKYANSE
DYTKQVSIHNNCSKTYNKGDYIVDLKRPVKNRMKYTWSAGLYTTQSAENETEDSGK
ASTKRSKRRHRRSDCMCAICVLKRRKREPRDSGAOEESPSPSDVSNVNM
GDGMDIWDKRPDEQEKTEIYELDEPVSKTOVLENKOEVEENVESENKRYANE
DKTQSDISEMETGEDEPYNASAEKILVYASLEGRSTONEEERKEKLOOKRLE
REMKRKKOEKFOVYNNPOLSLCETLPEPNNNHNSVWNGHSHLPRRGGNSRSALHK
AVESIMELNLSAISALGARSLPAMDSESGFHTLRLRAITSNASTIFITISKAIGEYASL
ISKAVELITFTDMRLILNKCKSEAKRRCODDAVAGGRSRADPEPGDMSSELLR
AAVSPVPSRRKPTPEEQRSQSLGVGAEIWSVPAELIIRSSSEPOCLVQSGDGA
RQNGCGEIVIAHMLCQVLLIRAPLCYLL"

repeat_region
19473..19512
/rpt_family="(A)n"
complement(24358..24296)
/rpt_family="AT-rich"
gene
complement(25863..36393)
/gene="T15M6.5"
/note="pseudogene, putative cationic amino acid transporter"
pseudo
complement(30166..30215)
/rpt_family="(A)n"
complement(32458..32485)
/rpt_family="AT-rich"
complement(32989..33008)
/rpt_family="(TA)n"
complement(33253..33276)
/rpt_family="(TA)n"
complement(33753..33773)
/rpt_family="AT-rich"
complement(33893..33928)
/rpt_family="AT-rich"
37880..38594
/gene="T15M6.20"
/note="pseudogene, putative retroelement pol polypeptide"
pseudo
complement(33920..39376)
/rpt_family="AT-rich"
complement(39499..39523)
/rpt_family="AT-rich"
39684..39709
/rpt_family="(TAA)n"
complement(40447..40527,40617..40739,40922..41017)
41217..41405,41490..41576,41787..41853,41997..42067,
42272..42409,42496..42597,42686..42769,42978..43034,
43117..43206,43297..43437,43824..43928,44069..44221,
44305..44408,44502..44607,44691..44874,44972..45057,
45293..45397,45493..45507,45745..45845,46042..46119,
46206..46328,46511..46587,46694..46882,46982..47093,
47463..47681,47870..48243,48829..48479,48788..48823,
48927..49132,49211..49316,49403..49696)"
/gene="T15M6.6"
complement(40447..49696)
/gene="T15M6.6"
/note="contains similarity to maleless protein (mle)
GI:157905 from (Drosophila melanogaster)." "
CDS
41217..41405,41490..41576,41787..41853,41997..42067,
42272..42409,42496..42597,42686..42769,42978..43034,
43117..43206,43297..43437,43824..43928,44069..44221,
44305..44408,44502..44607,44691..44874,44972..45057,
45293..45397,45493..45507,45745..45845,46042..46119,
46206..46328,46511..46587,46694..46882,46982..47093,
47463..47681,47870..48243,48829..48479,48788..48823,
48927..49132,49211..49316,49403..49696)"
/gene="T15M6.6"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAG50700.1"
/db_xref="GI:12321256"
/translation="MAPTKKPKKKKSKNSKNEIASLINSCKKKPKKPKLISPENED
RIRRLILNRKPSVYATVTSVTKRRKLNLYENLSCSEFPLNDQIETLVYLSRDAT

Query Match	1.5%	Score 98	DB 12	Length 103637
Best Local Similarity	49.2%	Pred 1.3e-06		
Matches 375	Conservative	0	Mismatches 375	Indels 12
			Gaps	4
QY 4496	tgagataaagtcgaatgctgaactagatgagaatattagatcacaagaatglttaccgaaga	4555		
Db 88790	TGGAGAAATCCATGGATGAGAGATCAAAATCATATACAAAGAAATGACACTGGAGATT	88731		
OY 4556	gttccaatgccacccggtgtgaagccatatactatlygggttgggtacactgagaat	4615		
Db 88730	ACTTCACCTTCCAAATGGACACAAAGCCAAATGGCGTGAAGTGGCTGTATAAAGCAAGAAA	88671		
OY 4616	gattctccaaggtgtgtctgc---gaatcacgttlytfttgcctgtgcaacaagaa	4672		
Db 88670	AACCTTAAGGAGAAAGTGGAAAGATACAAAGCAGACGATGGTTGTGCAAAAGTTATAGTCA	88611		
OY 4673	aaggaanaacttgattatgacctttagttagttagtaccctgttatagactctgtact	4732		
Db 88610	AGAGCCCGAATTTGACTATGAC-----GAGGTATTTTGCTCCCGTTGCGTCAAGAAAG	88557		
OY 4733	ataagattatgcaataataggttgtgtaattaggaatgacaatccaactttagacgtc	4792		
Db 88556	GTTATGACTATATCATCTCACTAGCGGCTCAAAACAAGTGAAGATACATCAATGGATGTC	88497		
OY 4793	gagtcgcgcatcaatcaatccctattactcaatcaatcaatcaatcaatcaatcaat	4852		
Db 88496	AAGTCGGCCTCTTCTTAATGAGATCTTGAGAGAAAGTTTACATTGAGCAACCAAGC	88437		
OY 4853	aaatcagtlaccttgaagaaaaaacatgtgtgtgtattgaaagcttgcctatgggtta	4912		
Db 88436	TACATAGTCAAGGTGAGAGAGCAAAAGTCTTGAGACTTAAAGAGCGCTTTATGGATT	88377		
OY 4913	aaacagcgcggttggaaaggtatcacactatcaaaagatgtggaagacatgtgtt	4972		
Db 88376	AAACAGACCCCAAGACCTTGGAT-ACCTGAAATTGACAAAGATTTCAAGAGAAAGATT	88318		
OY 4973	actcaagtttacaacaatgatgttatttccatctgataatgaagagggatcaatata	5032		
Db 88317	CATCAAGTGT--CCATATGAGCATGCGACTTATATCAAAATTTCAAAAAGAAATATATT	88260		
OY 5033	tatttagtttataatgttgaatatacttataatgttgaagtlccaaaaaagtattgat	5092		
Db 88259	ATCGCATGCTTATATGTAGATAGCTGTGATATTCACGGGTAACTCCAGCATGTTCGA	88200		
OY 5093	aatttggatcaatttggatgagatcaattttagaagtaagtttgggtgaatataaat	5152		
Db 88139	GAATTCACAGAAAGATGACGAAGGAGTTCCGATGACGGACATTTGATGATGCTTAC	88140		
OY 5153	tatttggatgaatgaatttgcataaacccgaatctgttatacttatactcacaanaatt	5212		
Db 88139	TATCTCGAATTTAGATTAACAAAGAGCATGGAATATTTCAATTAATCAAGAAAGGTTAT	88080		
OY 5213	ctcaagaacttactaaggatttcaaatagatgactcaat	5254		
Db 88079	GCTAAGGAGGTACTTAAGAAAGTTCAAGATGATGACTCAAAAT	88038		
RESULT 14				
AC007259/c				
LOCUS	AC007259	97146 bp	DNA	PLN
DEFINITION	Arabidopsis thaliana chromosome I BAC T28P6			17-AUG-1999
ACCESSION	AC007259			complete sequence.
VERSION	AC007259.4	GI:5306232		
KEYWORDS	HTG.			
SOURCE	thale cress.			
ORGANISM	Arabidopsis thaliana			
REFERENCE	Okuyama, Y. Viridiplantae: Embryophyta: Tracheophyta: Spermatophyta			
AUTHORS	Brassicaceae: Brassicaceae: Arabidopsis.			
	1 (bases 1 to 97146)			
	Federspiel, N.A., Palm, C.J., Conway, A.B., Conn, L., Hansen, N.F.,			

Altefi, H., Araujo, R., Huizar, L., Rowley, D., Buehler, E., Dunn, P., Gonzalez, A., Kremenetskaia, I., Kim, C., Lenz, C., Li, J., Liu, S., Luros, S., Schwartz, J., Shinn, P., Toriumi, M., Vyotskaia, V., Walker, M., Yu, G., Ecker, J., Theologis, A. and Davis, R.W.

Unpublished

2 (bases 1 to 97146)

Federispiel, N.A., Palm, C.J., Conway, A.B., Conn, L., Hansen, N.F., Altefi, H., Araujo, R., Huizar, L., Rowley, D., Buehler, E., Dunn, P., Gonzalez, A., Kremenetskaia, I., Kim, C., Lenz, C., Li, J., Liu, S., Luros, S., Schwartz, J., Shinn, P., Toriumi, M., Vyotskaia, V., Walker, M., Yu, G., Ecker, J., Theologis, A. and Davis, R.W.

Direct Submission

Submitted (08-APR-1999) DNA Sequencing and Technology Center, Stanford University, 855 California Avenue, Palo Alto, CA 94304, USA

3 (bases 1 to 97146)

Federispiel, N.A., Palm, C.J., Conway, A.B., Conn, L., Hansen, N.F., Altefi, H., Araujo, R., Huizar, L., Rowley, D., Buehler, E., Dunn, P., Gonzalez, A., Kremenetskaia, I., Kim, C., Lenz, C., Li, J., Liu, S., Luros, S., Schwartz, J., Shinn, P., Toriumi, M., Vyotskaia, V., Walker, M., Yu, G., Ecker, J., Theologis, A. and Davis, R.W.

Direct Submission

Submitted (01-JUL-1999) DNA Sequencing and Technology Center, Stanford University, 855 California Avenue, Palo Alto, CA 94304, USA

4 (bases 1 to 97146)

Federispiel, N.A., Palm, C.J., Conway, A.B., Conn, L., Hansen, N.F., Altefi, H., Araujo, R., Huizar, L., Rowley, D., Buehler, E., Dunn, P., Gonzalez, A., Kremenetskaia, I., Kim, C., Lenz, C., Li, J., Liu, S., Luros, S., Schwartz, J., Shinn, P., Toriumi, M., Vyotskaia, V., Walker, M., Yu, G., Ecker, J., Theologis, A. and Davis, R.W.

Direct Submission

Submitted (17-AUG-1999) DNA Sequencing and Technology Center, Stanford University, 855 California Avenue, Palo Alto, CA 94304, USA

On Jul 1, 1999 this sequence version replaced gi:5103851. Bases 90095-97146 of clone T28P6 overlap with bases 1-7052 of 'TAU', BAC clone T19D16, gb|ATU95973.

e-mail for correspondence: arabidopsis.stanford.edu

Genes with similarity to proteins in the databases are described as 'putative', 'like' or 'similar to'. Genes that have EST similarity but no significant protein similarity are described as 'unknown proteins'. Genes that are annotated based only on gene prediction software are described as 'hypothetical proteins'. The software programs used to predict genes include: Gail (Informatics Group, Oak Ridge National Laboratory), http://compbio.ornl.gov/section/index.html, GENSCAN (Chris Burge, http://genome.stanford.edu/chris/GENSCAN.html), Fexa (V Solovayev & A. Salamov, Sanger Centre, http://genome.sanger.ac.uk/), and NetPlantGene (S.M. Hebsgaard, et al., CBS, Technical University of Denmark, http://www.cbs.dtu.dk/NetPlantGene.html).

Location/Qualifiers

1. 97146

/organism="Arabidopsis thaliana"

/db_xref="taxon:3702"

/chromosome="1"

/clone="T28P6"

864. 3964

/gene="T28P6.1"

/note="RK51"

join(864..2139,2289..2408,2497..2670,2798..2917,3106..3343,3422..3566,3650..3964)

/gene="T28P6.1"

/note="identical to cDNA receptor-like protein kinase clone, g14008007"

/codon_start=1

/product="receptor-like protein kinase"

/protein_id="A04998.1"

/db_xref="GI:5734723"

/translation="MKVYVFIFFFLPFCISVDTIMRROSLRDGEYILSAGKRFAGFSLGDELRYGIMVQISQGITVWVANDHPINDISGKVSNNRNLVSVDNMT

ELIWTNSDMSLEPLVATLSDGLNLVLEFVTKGSPFESDHPDIDPELPMRLGFT
RKGLDIRLSITKMSGDPSPGSDILRMERGGPOLLYXGVPMRMGWTGKHSVG
PEMDIPIYINNSVNNEDNESTFYGTGDAVITRMVNTGTMGTWATARKRND
MSYKRCQCDNYALNCGPNGVCDSPSSSTFECTCLPGREPFPHMELRDSGGCTKKR
ASICSEKDGDMVTLKRMKIPDTSDASVDNMFLKQCRORLANSQVAASAHESKRG
AICGLKMGDMADARTYLNLSGDFYIRVKEELIARNNRNLGSKRRVLLIISLAAV
MLTIVILFCVRRERKSNRHSANPAMPDFPDEFEEDOKARNEPLDPLDNTI
VAATINNESSONKAGRAVIRKPYGDSGEVEYKLGNGVROGADIKVAKSHEE
ORLELMPRMEIRVIGIARIGIYIHODSRLRIIRHDKASNLILDSMTPKTSDRGA
RITGNGOMGCKSRVYGTGGYAPAMGORSIKSDYSPRGVILLETITTKGNAPH
EESNLVGHITWIDMENGERTIINDIMDETDEREYKTCITGILCVOENASDNDVM
SSVIMLGNATILNPKHPAFISARRRGENGACILKGTGTISVNDVTFSDIGR

complement(4664..7708)

/gene="T28P6.2"

complement(join(4664..4972,5054..5204,5275..5427,5661..5821,5908..6101,6189..6296,6391..7708))

/gene="T28P6.2"

/note="Very similar to receptor-like protein kinase"

/codon_start=1

/protein_id="A04998.1"

/db_xref="GI:5734724"

/translation="VVSVYIRRRFVLLIACCLISRLRCGEDRTFESSPIKDS
EPLTCSGIFRCEPFPMVSTRIRRVGTWYKPIQTYVWANKDSPINDSGYIST
YODGNLAVIDGRNRRLVSTNSVPAAPNATVQMLQDNRMNGLIMESFKH
PYDSFMRPTLGTGRTGNLKTWSHSDSTGNYTAGLAPTEPELLMKNNVPT
WRSGPNNGQVFIOLPMDSLFLDGLNLSMDQGTISMSYANDSMYHNPDLPEGIIY
OKWSTSMRTKRGVFPYTDGAYRCRPFESCHANGENPCKCYKNGKNTWNG
GDNSSCTMRAPLDCGRORNSVGGGGAGDFLKKKKVCIISERSASQVPCV
CLNNGSCYATVADRGICMLMSGDIYDMSGIDFLIRAHSEILKTHSLAWIA
APYIGMLIAACVILACRKYKRRPAKDRSAELMFKRMEALITDNSASQILAKE
LPLEFQVATATSDSFLSRNLKGQGGGVYVYGRKLEPGQELAVKRLSRKSGGLELM
NEVVYISKLQHRNLVLKCCIEGERIDSRILIRDKASNLILDENLKLKISDG
LARIFRANDEANTRRVGTGYCSPSEYMEGFESEKSPVSLGVIETISGRNNS
SHREENNLNLAVAKMLDNGEALSLADVADKPEKTECHVIGILCVOEVANDR
PNSVYIMLTETNMSLADPKQPAFIVRGADEALSSQSSQKINSVSLTAVGR

complement(9352..10839)

/gene="T28P6.3"

complement(join(9352..10047,10150..10617,10729..10839))

/gene="T28P6.3"

/note="Hypothetical protein"

/codon_start=1

/protein_id="A04999.1"

/db_xref="GI:5734725"

/translation="MDPLASVLEKYSKFTSSQDLYSRHGHENPSRLNPIIDIKRL
QRTFSDMLKLEKROEKYRIISQKLSGGFQSTSVHREVVLAAILMGNTDE
ESFNGLNKGEVGRGLSRVFEFSLRETDRLVAELVAGVSGHRDLGGELSLAKY
FYKADINDMESAIAIPVGAHFDIASVSYQGSLEFVSELGPLSOHNGSAIGL
TFKSNITRTSLAOSMSNVNEGGLAPAPNCPFTPGOVCHILRSKLILGCHOVLT
SSNHSVGAITVYPSLIRPRTTPEPSPAPLPMRNMHVSSSIAIKLDSIMEST
KIGKGTETONSREKOVKMSVSTTNDPEDEVGMSVGGVDSGRHDOFOVESYLYFN
IDDRFLSPCLVYHNSDGRITGLMKSHMS"

14210..17181

/gene="T28P6.4"

join(14210..14332,14521..14578,14694..14959,15053..15169,15263..15323,15394..15557,15641..15735,15821..15871,15958..15998,16078..16251,16342..16377,16463..16530,16620..16680,16774..17181)

/gene="T28P6.4"

/note="Highly similar to MLO proteins: Highly similar to barley MLO (control element of plant pathogen resistance) multi-gene family in arabidopsis"

/codon_start=1

/protein_id="A04999.1"

/db_xref="GI:5734726"

/translation="MADQKERTLEETSTWAAVAVFVILFSTVLEHSITKIGTWK
KHKRALPALKKVAEMLLGLFISLITIGTPTSNICISQKASTHPPSAABEAK
KTKGDAGKKGDDGDKRGLLELASVYIHRSLATKGYDKCAEKQVAFVSAYGI
HQLHIFLVLAVVAVVYICVYAFKIKMRKSWEEETKYTEQVSNDRPFRFARD
TSFGRHNLNFMKSTRVTLVIVCFRPGFSGVYKVVYLRLRGFIIMAHNAPDNESFDF
RKYIORSLEKDKETVETSPVIMFAVLEFLNLSGYSMLPELPIVLVILGTLK
EYITIKRIGRDERKGDVYAGVAPVGGDLPMPGKPRILILHLVIFLTNFOALAPA
MSTYEPNLNCHNESTADVITRLVYGAVALICSTVYTLPLATLYOMGSKKPKFYEND

gene	complement(17361..20311) /gene="T28P6.5" complement(join(17361..17669,17749..17878,17963..18200,18244..18494,18566..18744,18826..18933,19024..20311)) /gene="T28P6.5" /note="Very similar to receptor-like protein kinases" /codon_start=1 /protein_id="AADA9992.1" /dx_xref="GI:5734727" /translation="MRLESISPVHVLSDFPLSVSLAHERALFSGTINDSETIVS SFRFRGFEFSPVNSTNRVGIWNSIPVQVIVWAKDPIINDSSGIVSIEGGINLV VTDGQRLWSTNSTSRASNTVAELLESALVLAIPYDPLFENNDDNATWREGS MVLTNRATGGGNTITTSWNPSPDSPGSYALVLAIPYDPLFENNDDNATWREGS PWNINMGILPDVYPGILFVFRKNDQNPGRATSNVSDTLRLHYLDYRFAIRPDM SEARNMTLSQVYATPCDIYSRCGQTTCCPKRNPHCSTCKGRPNLLDMNNGNMS GGCTRKPLQECRYONNNGSADRELTLMKMPKPDPAKRSSEAPCEPMTQSCSIAF AHGIGYGCMTWNRSLVDSOVLVLSASGMNDISLHASEKTDQDRPILIGTSLAGIIPV ATCVLLARIVYMKRRAKKGTDAEOLIKRVEALAGSGREKLELPDEFVYLAATNP FESLNLGGGGFCEVYMGKLEGGDEIVKRLSGOGGLEELVEVVYISKQLRHLV KLFECCIJAGEERMLVYEFMPKSLDFIPIPREKLLDMNTREIINGICGLIHRIV DSRLRIIHRDLKASNIILDEMLPKISDFGARLPENGDENRRVRYGYAPRY AMGLESEKSDVPSGLVILLETISGRNHSHTLLAHWSIYNBSEINGMDPELFDQ PEKTRKCVHIALCVDAAANDRPSVTCMWSSEVADIPEKQPAFMPRNVLEAE FSESIALKASINNTVIVDSGR" complement(20734..23691) /gene="T28P6.6" complement(join(20734..21042,21122..21251,21332..21569,21652..21862,21947..22125,22207..22314,22404..23691)) /gene="T28P6.6" /note="Very similar to receptor-like protein kinases" /codon_start=1 /protein_id="AADA9993.1" /dx_xref="GI:5734728" /translation="MRLESSEPVCLIVLSCEPLSVSLAORAFSPSGKINDSETIVS SFRFRGFEFSPVNSTNRVGIWNSIPVQVIVWAKDPIINDSSGIVSIEGGINLV VTDGQRLWSTNSTSRASNTVAELLESALVLAIPYDPLFENNDDNATWREGS MVLTNRATGGGNTITTSWNPSPDSPGSYALVLAIPYDPLFENNDDNATWREGS PWNINMGILPDVYPGILFVFRKNDQNPGRATSNVSDTLRLHYLDYRFAIRPDM SEARNMTLSQVYATPCDIYSRCGQTTCCPKRNPHCSTCKGRPNLLDMNNGNMS GGCTRKPLQECRYONNNGSADRELTLMKMPKPDPAKRSSEAPCEPMTQSCSIAF AHGIGYGCMTWNRSLVDSOVLVLSASGMNDISLHASEKTDQDRPILIGTSLAGIIPV ATCVLLARIVYMKRRAKKGTDAEOLIKRVEALAGSGREKLELPDEFVYLAATNP FESLNLGGGGFCEVYMGKLEGGDEIVKRLSGOGGLEELVEVVYISKQLRHLV KLFECCIJAGEERMLVYEFMPKSLDFIPIPREKLLDMNTREIINGICGLIHRIV DSRLRIIHRDLKASNIILDEMLPKISDFGARLPENGDENRRVRYGYAPRY AMGLESEKSDVPSGLVILLETISGRNHSHTLLAHWSIYNBSEINGMDPELFDQ PEKTRKCVHIALCVDAAANDRPSVTCMWSSEVADIPEKQPAFMPRNVLEAE FSESIALKASINNTVIVDSGR"
CDS	
gene	
Query Match	1.5%: Score 97.2; DB 12; Length 97146;
Best Local Similarity	47.8%: Pred. No. 1.86-06;
Matches 375; Conservative	0; Mismatches 403; Indels 6; Gaps 3;
OY 4471	caaatgactataatgatccgggaatgataaagtcaatgaatgctgaactagagaat 4530
Db 34323	CCAAAACCCCTTGAAGAAAGAAAGACTTTGGAGAAATGCCATGTAGTAAGATCAAAATCAAT 34264
OY 4531	tagatcaaaagaatgttttcgaagaagttccaatcccccoggtggaagccataatcat 4590
Db 34263	ACAAAAGATATACATGGGAGTTAATTCTCACTTCCAAATATGACACAAAGGCATTTGGCGCT 34204
OY 4501	ggattgggtacatactcgaagaatgatcttcctccaagaagtgttctcggaaatcacgttg 4650
Db 34203	GAAGTGGGTGTATTAAGCAAGAAAGAAACTCT---AAAGACGAAGTGAAGATACAAAGC 34147
OY 4601	tgttgcatatgacgacagacacaaagaanaaatgtgatatgaacctttagtgtatgtc 4710
Db 34146	AAGATTGGTTCCAAAGGTTTATGTCAAAAGATGCGAATTAATATGACGAGGATTTTGGC 34087
OY 4711	acctgtatagatctctgacatataagatttgcaataatagtttgtaattagaat 4770
Db 34086	TCCCGTTCCTGCTGTAGAAAGGTTAGACTATATCTATCTACTAGCGGCTCAAAACAAGTG 34027
OY 4771	gacaattcaacatttagacgtcgagtcggcgatctataagtcctctattactcatc 4830
Db 34026	GAAGATACATCAAAATGATGTCAAGTCCGCGCTTTTAATATGAGACGATTTGAAGAAGAGT 33967
OY 4831	tccaattatgtcttctcctaataatgtaacttcttgagaanaaacatttgttgattc 4890
Db 33966	TTACATTATGACACACAGGCTCATATATGTCAAAAGTAAAGAAGCAAAAGCTTTGAGGCT 33907

QY	4891	gaaacgtctgtcctaagggttaaacacagtcgggttggaatggtatcaacatcaaaag	4950
Db	33906	AAAAAGGTGCTTTATGGATTTAAACCAAGCCCAAGAGCTTGGAAAT-ACTCGAATTGGACA	33848
QY	4951	agtaattgaagacattggttttactcaaggttttacacaatgaatggttttattacattga	5010
Db	33847	AGTATTTCAAGGAGCAAAATTTTCAAGTGTG-CCATATGAGCAGTACCTTATATCA	33790
QY	5011	atataaaggagatcaagtaataattatggtttatataatgtttgatgtatcttaagttgg	5070
Db	33789	AATTCAAAAGAGATATATTTGATGCGATGCTTATATGTAGATGCTGATATTCACGGG	33730
QY	5071	aagttcacaaaaagattatgttaattctgtgatcaattgaagaagatcaatttgaagttaa	5130
Db	33729	TAAACAATCCAGCATATTTTCGAAGATTTCAAGAAACAGATGACGAAGAGATTGCAATGAC	33670
QY	5131	agtgcttggatgaataataatcatctgtgatgtgaatttcgttaaacccaatctgtgta	5190
Db	33669	GGACATTTGATTTGATGCTTACTATCTGCGAATTTGAAATTAACAAAGAAAGCAATGGAT	33610
QY	5191	tattttactcaagaanaattttccaaagaataatttaagatttcaacatagatgattc	5250
Db	33609	ATTCTATTACTCAAGAGGCTATGCTAAGGAGGTACTTAAAGAGTTCAAGATGATGATGACTC	33550
QY	5251	atat	5254
Db	33549	AAAT	33546
RESULT	15		
ATLECRECK			
LOCUS	ATLECRECK	4793 bp	DNA
DEFINITION	A. thaliana	lectin receptor kinase gene.	PLN
ACCESSION	Y08010		
VERSION	Y08010.1	GI:1769896	
KEYWORDS	Art1; copia-like element; lectin receptor kinase; transposable element.		
SOURCE	thale cress.		
ORGANISM	Arabidopsis thaliana		
REFERENCE	Eukaryotes: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis. 1 (bases 1 to 4793)		
AUTHORS	Herve,C., Serres,J., Dabos,P. and Lescure,B.		
TITLE	Nucleotide sequence of DNA art1 (accession no. Y08010), a new Arabidopsis copia-like transposable element, disrupts a lectin receptor kinase gene in the Columbia ecotype (PGR97-059)		
JOURNAL	Plant Physiol. 113, 1464-1464 (1997)		
REFERENCE	2 (bases 1 to 4793)		
AUTHORS	Herve,C.		
TITLE	Direct Submission		
JOURNAL	Submitted (13-SEP-1996) C. Herve, INRA-CNRS, LBMRPM, Chemin De Borde Rouge Bp27, Castanet-Tolosan Cedex, 31326, FRANCE		
COMMENT	Overlaps with Y08020.		
FEATURES	Location/Qualifiers		
source	1..4793		
	/organism="Arabidopsis thaliana"		
	/sub_species="Columbia"		
	/db_xref="taxon:3702"		
	/clone_lib="J68"		
	/lab_host="EMBL-3 CLONTECH"		
	1..439		
	/note="Art1 transposable element"		
	<423..>2054		
	/codon_start=1		
	/product="lectin receptor kinase"		
	/protein_id="CA69271.1"		
	/db_xref="GI:1769897"		
	/translation="KNFTLIRKWSLILKMSNNPVPVLTKSNVDMKSLPMKAL		
	LGADWEIETGRIPEENGCSISQPKDILRSKRDKKALCIYGDIEDDPFKYV		
	EAHSKAKMEKLRSTYKGAQVKKVRLQTLRGFEALQMRDGEIYSYFSSRLVTNN		
	LKRNGEALDVRIMEKVLRLDLKFEHIVIVIEETKDLQEMTIEQLGSLQAYEEKRK		

